

# CONSULTANT REPORT

## ANALYSIS OF PUBLICLY OWNED UTILITY REPORTED AND VERIFIED SAVINGS

Review of Energy Efficiency Program  
Savings Estimations in Annual Reports and  
Measurement and Evaluation Studies

Prepared for: California Energy Commission

Prepared by: KEMA, Inc



NOVEMBER 2010

CEC-200-2010-008

***Prepared by:***

***Primary Author(s):***

Betty Seto  
Eric Swan  
Fred Coito

***Contributing Author(s):***

Karin Corfee  
Carrie Webber  
Sam Golding  
Jonathan Taffel

KEMA, Inc.  
155 Grand Street  
Oakland, CA 95612

***Contract Number: 400-07-032***

***Prepared for:***

**California Energy Commission**

Che McFarlin  
***Project and Contract Manager***

Bill Junker  
***Office Manager***  
***Electricity Supply Analysis Division***

Sylvia Bender  
***Deputy Director***  
***Electricity Supply Analysis Division***

Melissa Jones  
***Executive Director***

**DISCLAIMER**

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees or the State of California. The Energy Commission, the State of California, its employees, contractors and subcontractors make no warrant, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

## ABSTRACT

This report summarizes the content and scope of evaluation, measurement, and verification studies submitted by publicly owned utilities, in accordance with Assembly Bill 2021 (Levine, Chapter 734, Statutes of 2006) and Senate Bill 1037 (Kehoe, Chapter 366, Statutes 2005). These statutes require the installation of all cost-effective energy efficiency measures and an independent assessment of reported savings.

The key objectives of this study are to:

- Review publicly owned utility savings estimation in annual reports and evaluation reports.
- Develop and use a framework to analyze evaluation reports.
- Compare reporting and evaluation requirements of publicly owned utility programs with investor-owned utility programs.

As of March 2010, 11 publicly owned utilities have submitted 12 evaluation reports. The framework developed by KEMA consists of 21 criteria across the following 5 categories: contextual reporting adequacy, overview of the specific evaluation report, gross savings, net savings, and summary/conclusion of the evaluation report.

The evaluation reports submitted by publicly owned utilities to date represent a first effort to independently verify savings associated with program measures. Many of the publicly owned utilities have had little to no experience with program evaluation before these efforts. It remains a challenge to reconcile the verified savings in the evaluation reports with the reported savings under Senate Bill 1037. Publicly owned utilities appear to have invested minimal funding for evaluation, measurement, and verification.

Final recommendations include establishing mandatory components to be included in future evaluation reports, for example program, description and associated reported savings. The Senate Bill 1037 annual report submitted by the publicly owned utilities should include savings by program. Finally, the results suggest that more education and training on the expected rigor, activities, and scope of the evaluation efforts will improve the reliability of the verified savings.

**Keywords:** Verified savings, energy efficiency, program, evaluation, publicly owned utility, investor owned utility, ex ante, ex post, independent third party, Senate Bill 1037, Assembly Bill 2021

Please use the following citation for this report:

Seto, Betty, Eric Swan, Fred Coito. (KEMA, Inc.). 2010. *Review of Energy Efficiency Program Savings Estimations in Annual Reports and Measurement and Evaluation Studies*. California Energy Commission. Publication number: CEC-200-2010-008.

## TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>i</b>
<b>TABLE OF CONTENTS.....</b>	<b>ii</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>1</b>
<b>CHAPTER 1: Introduction.....</b>	<b>6</b>
Study Scope and Key Objectives.....	7
Framework to Evaluate EM&V Studies.....	7
Report Organization .....	9
<b>CHAPTER 2: Contextual Reporting Adequacy .....</b>	<b>10</b>
Portfolio Level Reporting.....	11
Comprehensiveness of EM&V Efforts .....	13
<b>CHAPTER 3: Assessment of Measurement and Evaluation Studies .....</b>	<b>17</b>
EM&V Studies Submitted and Reviewed.....	17
Scope and Rigor of EM&V Reports Submitted.....	19
Measure Life and Lifecycle Savings .....	22
Documentation .....	23
Gross Savings Approaches .....	25
Verification.....	26
Gross Savings Methods .....	27
Baselines .....	29
Sampling.....	30
Execution.....	34
Expansion of Sample to the Population.....	37
Discussion .....	39
Net-to-Gross.....	41
Sampling and Precision.....	43
Net-to-Gross Methodology.....	44
EM&V Summary and Conclusions.....	47
<b>CHAPTER 4: Comparison of POU and IOU Evaluations .....</b>	<b>49</b>

Budget Considerations .....	49
Realization Rates and Net-to-Gross Ratios.....	51
Reporting Requirements .....	56
<b>CHAPTER 5: Conclusion .....</b>	<b>60</b>
Key Findings and Recommendations .....	60
Timing of EM&V and SB 1037 Reporting Requirements .....	60
SB 1037 Documentation Adequacy .....	61
Contextual Reporting Adequacy .....	61
Gross Savings Approach.....	62
Sampling Approach.....	63
Net-to-Gross Analysis .....	64
Guidance for POU Evaluation Efforts.....	64
EM&V Reporting Requirements and Documentation.....	65
Conclusion .....	66

## **APPENDIX A: Summary of Each EM&V Study Assessed**

## **APPENDIX B: Example Guidance for POUs**

## **TABLES**

Table 1. Summary of POU Evaluation Efforts Submitted .....	10
Table 2. Summary of EM&V Studies Submitted .....	17
Table 3. Target and Achieved Precision by Study .....	33
Table 4. Required Protocols for Participant Net Impact Evaluation.....	41
Table 5. POU Energy Efficiency and EM&V Budgets.....	50
Table 6. Summary of POU EM&V Realization Rates.....	51
Table 7 Key Parameters from CPUC Evaluation Reports for 2006-2008 IOU Programs .....	52

## FIGURES

Figure 1. Criterion 1.1 – Does the EM&V study report the same portfolio savings values as the SB 1037 annual report? .....	12
Figure 2. Criterion 1.2 – Approximate percent of claimed SB 1037 savings covered by the POU evaluations. ....	13
Figure 3. Criterion 1.3 – Discussion of uncertainty related to the programs and claimed savings not evaluated. ....	15
Figure 4. Criterion 2.1 -- What is the overall approach of the EM&V report – does it take a program by program approach, or does it try to evaluate the whole portfolio (on a measure category or customer sector basis)? .....	19
Figure 5. Criterion 2.2 – What level of rigor was the evaluation generally conducted at?.....	20
Figure 6. Criterion 2.3 – Does the evaluation include an assessment of EUL and lifecycle savings? .....	22
Figure 7. Criterion 2.4 – Does the evaluation report provide sufficient documentation for a complete review? .....	23
Figure 8. Criterion 3.1 – Does the EM&V report include a suitable measure installation verification?.....	26
Figure 9. Criterion 3.2 – What was the gross savings evaluation method utilized?.....	27
Figure 10. Criterion 3.3. Is the choice of baseline suitable? .....	29
Figure 11. Criterion 3.4. Is the sampling approach appropriate?.....	30
Figure 12. Criterion 3.5 – Were the sampling precision targets stated and the achieved precision of the estimates reported? .....	31
Figure 13. Criterion 3.6. Was the selected savings approach executed appropriately? .....	34
Figure 14. Criterion 3.7 – Are the results extrapolated to the program population, in an adequate way? .....	37
Figure 15. Criterion 3.8 – Does the report properly explain difference between ex ante and ex post savings estimates? .....	39
Figure 16. Criterion 4.1 – Does the evaluation include a quantitative assessment of net-to-gross? .....	42
Figure 17. Criterion 4.2. Is the sampling approach appropriate?.....	43
Figure 18. Criterion 4.3 – Is the selected NTG analysis method applied appropriately? .....	44
Figure 19. Criterion 4.4 – Does the approach account for partial free-ridership?.....	46

Figure 20. Criterion 5.1 – Does the EM&V report provide clear recommendations for improving program process to improve realization rates and/or verified kWh results? .....	47
Figure 21. Example Checklist for POU EM&V report .....	66
Figure 22. Example of the “Results” tab of the 2010 EE Reporting Tool.....	B-1





## EXECUTIVE SUMMARY

Electric and gas utilities in California are required by Assembly Bill 2021 (Levine, Chapter 734, Statutes of 2006) and Senate Bill 1037 (Kehoe, Chapter 366, Statutes 2005)<sup>1</sup> to install all cost effective and feasible energy efficiency measures. SB 1037 specifically requires all publicly owned utilities (POUs) to report to the California Energy Commission and their local governing boards about current and projected energy efficiency programs, including expenditures and savings. AB 2021 reaffirms SB 1037 mandates and requires publicly owned utilities to develop energy efficiency targets on a triennial basis and provide an independent third-party verification of their annual claimed savings.

The Energy Commission seeks to ensure that the independent evaluation, measurement and verification (EM&V) efforts of publicly owned utilities reported energy efficiency savings is thorough and transparent. The two primary purposes for conducting EM&V studies of energy programs in California are: to reliably document program effects, and to make program designs and operations more cost-effective at obtaining energy resources.

While the POUs have submitted SB 1037 annual reports since 2006, most POUs had not previously conducted evaluation studies. In 2009, 11 utilities submitted 13 EM&V studies of their 2008-2009 efficiency programs: Alameda Municipal Power, Lodi Electric Utility, city of Lompoc, city of Palo Alto Utilities, Port of Oakland, Redding Electric Utility, Roseville Electric, Silicon Valley Power, Sacramento Municipal Utility District, Truckee-Donner Utility District, and Turlock Irrigation District. This report examines these studies and makes recommendations to improve POU's evaluation and reporting requirements to promote transparency and rigor of estimated savings.

The results of this report are meant to assist the Energy Commission in assessing the reliability of claimed energy savings associated with POUs' energy efficiency programs. The key objectives of this study are to:

- Evaluate POU savings estimation in annual reports and EM&V studies.
- Develop and use a framework to analyze the 2009 and future EM&V studies.

An EM&V assessment framework was developed to provide a consistent and systematic approach to assess the EM&V studies performed for the California's POUs. The framework developed for this study is based on the fundamental components of EM&V studies, as defined by the 2006 California Evaluators' Protocols and associated 2004 California Evaluation Framework used by California's investor-owned utilities. It consists of 21 criteria in 5 different categories to assess the POUs EM&V studies.

---

<sup>1</sup> Text of bill available at: [http://info.sen.ca.gov/pub/05-06/bill/asm/ab\\_2001-2050/ab\\_2021\\_cfa\\_20060421\\_125425\\_asm\\_comm.html](http://info.sen.ca.gov/pub/05-06/bill/asm/ab_2001-2050/ab_2021_cfa_20060421_125425_asm_comm.html)

- **Contextual reporting adequacy** – Whether sufficient information is provided in the evaluation report to understand the program being evaluated and any ex-ante<sup>2</sup> claimed savings.
- **Overview of the specific evaluation, measurement and verification report** – Examines overall scope of the evaluation report, including programs evaluated, rigor level, lifetime savings, and sufficient documentation.
- **Gross savings** – Assesses the approach to estimating savings associated with the quantity installed and methods for verifying actual installation rates and improving engineering calculations.
- **Net savings** – How well the evaluation assessed the effect of the program and whether the measures would have been installed in the absence of the program.
- **Summary and conclusions** – Whether the evaluation provides sufficient recommendations for program improvements and areas where the evaluation itself could be improved.

The goal of the evaluation efforts should be to provide the Energy Commission with an independent assessment across the POU's portfolio of programs of total claimed savings. With exception of some of the largest utilities, most of the POU's have had little or no experience with program evaluation before these efforts.

## **Key Findings and Recommendations**

The report summarizes key findings and results of the analysis and application of the framework to evaluate EM&V studies. Recommendations are provided to improve the rigor, documentation, and consistency EM&V efforts to support SB 1037 and AB 2021 requirements. Suggestions for improvements are made realizing that they may not be practical or cost-effective for some of the smaller utilities.

### *Timing of Evaluation, Measurement, and Verification and SB 1037 Reporting Requirements*

There needs to be a clear relationship between the EM&V studies and the SB 1037 annual reports. For the most part, it appears that POU's are using the SB 1037 annual report to present the claimed (ex-ante) portfolio savings based on the EE Reporting Tool which they use to develop their annual savings and are not presenting the evaluated (ex-post) savings. Due to the range of program year completion dates, it is not feasible for all POU's to report ex-post<sup>3</sup> savings within the March 15 SB 1037 annual reports. Therefore, it is recommended that a regular reporting deadline for EM&V reports and verified savings be established.

---

<sup>2</sup> Ex ante is Latin for "beforehand." In models where there is uncertainty that is resolved during the course of events, the ex ante values are those that are calculated before evaluation, measurement, and verification studies have been conducted.

<sup>3</sup> Ex Post is Latin for "after the fact". In models where there is uncertainty that is resolved during the course of events, the ex post values are those that are calculated after evaluation, measurement, and verification studies have been conducted.

### *SB 1037 Documentation Adequacy*

The SB 1037 annual reports submitted for 2006, 2008, and 2009 only include net savings values. The SB 1037 report should specify both gross and net savings, to promote transparency in the net-to-gross assumptions. There should be a clear indication of whether reported savings are claimed (ex-ante) or whether evaluated (ex-post) savings are being reported. The Energy Commission should also request the completed EE Reporting Tool from the POUs as documentation of their SB 1037 claimed (ex-ante) savings, as well as calculations supporting the revision of default values in the EE Reporting Tool or addition of any custom measures.

### *Contextual Reporting Adequacy*

In general, the POU evaluation studies focused the greater part of their evaluations on the one or two programs that were responsible for the majority of portfolio savings, but the exact percentage of the total portfolio being evaluated was usually not reported. Information was lacking in the report to understand fully the savings being evaluated and how the results compare with the POUs claimed ex-ante savings at the portfolio level. This context is needed, especially to understand what the verified portfolio level savings would be. A summary report should be included that provides the portfolio level verified savings, indicating which measures savings or programs were not evaluated.

### *Gross Savings Approach*

The evaluations were heavy on installation verification with a review and critique of the deemed savings<sup>4</sup> method. Primarily this was used for residential appliance rebates and non-residential lighting retrofits. For appliance rebates, the verification and deemed savings review is appropriate. For other non-residential projects, however, this is generally not sufficient. This is especially true when these projects account for a large percentage of the POU savings portfolios and when most of these programs have not been evaluated previously. In most cases, the method and algorithms used to calculate evaluated gross savings and verification results were not clearly documented. The pathway from raw data to the final results was not clear in most cases, including baseline assumptions for the energy savings estimate. In the effort to improve the quality and transparency of evaluated gross savings estimates, the POU evaluations should assess a significant share of their reported savings at the basic level of rigor, provide all ex-ante assumptions, and document data collection tools.

### *Sampling Approach*

Many studies fell short of executing fully appropriate statistical methods. When a proper sample design is used, the results of sampled projects can be used to develop estimates of program population results that will be close to the true values that would have resulted if the

---

<sup>4</sup> Deemed savings are the estimated energy savings that are expected to be achieved for different energy efficiency measures. The primary source of estimated energy savings is California's Database for Energy Efficiency Resources (DEER). <http://eega.cpuc.ca.gov/deer>.

same data collection had been carried out for all projects. Many of the EM&V studies did not expand the evaluation results of the sample back to the program population at the program level. A specified confidence interval and desired relative precision should be used to determine the sample size necessary to meet those targets. EM&V studies should document both the desired relative precision and the achieved precision. Ultimately, evaluations should provide program level impacts resulting from the evaluation, measurement, and verification work.

### *Net-to-Gross Analysis*

The determination of gross savings indicates what savings take place while net savings shows what percentage of program savings occurred specifically due to the program. Although many of the EM&V studies mentioned net-to-gross, only a few studies developed net-to-gross ratios to adjust for net savings. Net-to-gross methods exist, including standardized self-report survey questions and analysis. The POU could adopt this approach for their EM&V studies. At minimum, the evaluation, measurement, and reports should either complete an adjusted net-to-gross ratio analysis for the program, or clearly indicate a “pass-through” of the deemed net-to-gross ratio is being used to estimate net savings.

### *Guidance for POU Evaluation Efforts*

Overall, it appears that POU evaluation funding is on the low side and may need to be increased if EM&V activities are to keep pace. Some POU might consider combining EM&V activities to gain some economies of scale, with this being especially appropriate for POU with similar efficiency programs. POU staff needs more guidance and training on the expected rigor, activities, and scope of the evaluation efforts to ensure the reliability of the verified ex-post savings. The Energy Commission may consider providing workshops, training tools, and webinars in EM&V principles, practices, budget setting and more specific topics.

### *Evaluation, Measurement, and Verification Reporting Requirements and Documentation*

To improve transparency and consistency of evaluation, measurement, and verification reports the POU should provide more documentation associated with evaluation efforts. This information would be useful to assess how effectively evaluation dollars are being spent and whether additional funds would improve the quality of evaluations. The POU may also benefit from a consistent reporting template for EM&V studies.

## **Conclusion**

Most POU are relatively new at EM&V studies for their efficiency programs and this may explain the results of this review of their initial efforts. The purposes of EM&V studies are to estimate the actual savings of a program and improve program delivery and savings value. This report and its framework of evaluation criteria are provided so that POU can use it as a guide to make practical and cost-effective modifications in EM&V practices for their energy efficiency programs.



# CHAPTER 1:

## Introduction

Electric and gas utilities in California are required by Senate Bill (SB) 1037 (Kehoe, Chapter 366, Statutes of 2005) and Assembly Bill (AB) 2021 (Levine, Chapter 734, Statutes of 2006)<sup>5</sup> to install all cost effective and feasible energy efficiency measures. SB 1037 specifically requires all publicly owned utilities (POUs) to report to the California Energy Commission and its local governing boards about current and projected energy efficiency programs, including expenditures and savings. AB 2021 reaffirms SB 1037 mandates and requires publicly owned utilities to develop energy efficiency targets on a triennial basis and provide an independent assessment of measured savings.

To meet the regulatory requirements, the California Municipal Utilities Association (CMUA) has developed a common reporting template to facilitate comparisons and compilations of the POUs' efficiency information, and facilitates the Energy Commission's statewide analysis since it is largely consistent with the investor-owned utilities' (IOU) reports.

The Energy Commission seeks to work with the POUs to ensure that the methods are thorough, transparent, and comparable to those used by the California Public Utilities Commission (CPUC) to evaluate the investor-owned utilities' portfolios. Consistent, robust, and independent evaluation is critical to ensure that energy efficiency can be depended upon as a resource.

There are two primary purposes for conducting evaluations of energy programs in California. These are:

- To reliably document program effects.
- To improve program designs and operations to be more cost-effective at obtaining energy resources.

While the POUs have submitted SB 1037 annual reports<sup>6</sup> to the Energy Commission since 2006, evaluation studies had not previously been completed, except for three by the Sacramento Municipal Utilities District (SMUD). In 2009, POUs completed 10 additional evaluation studies of their efficiency programs.

This report examines POU savings estimates in SB 1037 and the evaluation, measurement, and verification (EM&V) studies submitted before the end of 2009. Recommendations are included to improve POU evaluation and reporting requirements to promote transparency and rigor of estimated savings.

---

<sup>5</sup> Text of bill available at: [http://info.sen.ca.gov/pub/05-06/bill/asm/ab\\_2001-2050/ab\\_2021\\_cfa\\_20060421\\_125425\\_asm\\_comm.html](http://info.sen.ca.gov/pub/05-06/bill/asm/ab_2001-2050/ab_2021_cfa_20060421_125425_asm_comm.html)

<sup>6</sup> California Municipal Utilities Association, *Energy Efficiency in California's Public Power Sector: A Status Report*, December 2006, March 2008, 2009, and 2010.

## Study Scope and Key Objectives

The results of this study are meant to assist the Energy Commission in assessing the reliability of claimed energy savings associated with POU energy efficiency programs. This report examines several components of the POU savings estimates reported to the Energy Commission, including the development of the savings for the SB 1037 annual reports and the third-party verification of these claimed savings.

The key objectives of this study are to:

- Evaluate POU savings estimation in annual reports and EM&V studies.
- Develop and use a framework to analyze EM&V studies.
- Compare POU and IOU reporting and EM&V requirements.

This report summarizes the POU evaluation and measurement studies completed to date. An EM&V assessment framework was developed to provide a consistent and systematic approach to assess the EM&V studies completed for the California POU.

Recognizing the POUs have not received much specific guidance related to evaluation efforts, this study also examines IOU evaluation experience to date, and lessons learned that may be relevant to the POU efforts. In particular, the study compares POU and IOU requirements, including EM&V spending requirements, reporting requirements and EM&V rigor requirements.

## Framework to Evaluate EM&V Studies

The framework developed for this study is based on the fundamental components of EM&V, as defined by the *2006 California Evaluators' Protocols* and associated *2004 California Evaluation Framework*.<sup>7</sup> The CPUC developed these documents to provide valuable information concerning when evaluations should be conducted, the types of evaluation that can be conducted, and a discussion of approaches for conducting EM&V studies. These resources provide a rigorous systems approach to conducting evaluations so that all programs are able to document their effects and be compared to other programs and energy supply options.

The research team incorporated components of *the 2006 Evaluators' Protocols* and *2004 Evaluation Framework* to provide guidance to the POUs on methods and approaches that can

---

<sup>7</sup> *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals* ("Evaluators' Protocol"). Prepared for the California Public Utilities Commission. April

2006. [http://www.calmac.org/events/EvaluatorsProtocols\\_Final\\_AdoptedviaRuling\\_06-19-2006.pdf](http://www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf)

*The California Evaluation Framework*. Prepared for the California Public Utilities Commission and the Project Advisory Group. June

2004. [http://www.calmac.org/publications/California\\_Evaluation\\_Framework\\_June\\_2004.pdf](http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf)

lead to high quality evaluation studies. The framework developed by KEMA in 2010 to assess the POU EM&V studies consists of 21 criteria across the following 5 categories:

- **Contextual reporting adequacy** – Whether sufficient information is provided in the evaluation report to understand the program being evaluated and any ex ante<sup>8</sup> claimed savings.
- **Overview of the specific EM&V report** – Examines overall scope of the evaluation report, including programs evaluated, rigor level, lifetime savings, and sufficient documentation.
- **Gross savings** – Assesses the approach to estimating savings associated with the quantity installed and methods for verifying actual installation rates and improving engineering calculations.
- **Net savings** – How well the evaluation assessed the effect of the program and whether the measures would have been installed in the absence of the program.
- **Summary and conclusions** – Whether the evaluation provides sufficient recommendations for program improvements and areas where the evaluation itself could be improved.

The principal focus of the framework is to assess whether the evaluation efforts are producing reliable third-party verification of the POU claimed energy savings. The framework also reviews whether POU efficiency program effects are reasonably estimated in a defensible manner, reliably documented, and consistent with the evaluations completed for California IOU efficiency programs. Components of the framework also address cost-effectiveness and evaluation of program processes.

In determining the appropriateness of POU evaluation efforts, the framework uses the basic level of rigor, defined in the *recent 2006 California Evaluators' Protocols* and associated *2004 California Evaluation Framework*. Rigor is defined as the level of expected reliability. Higher levels of rigor lead to greater confidence that the results of the evaluation are both accurate and precise. The basic level of rigor represents the minimum requirements for evaluation. An enhanced level of rigor may be more appropriate for measures (or projects) contributing a large portion of portfolio savings, or with significant uncertainty associated with savings.<sup>9</sup>

By applying the framework to each of the evaluation reports submitted to the Energy Commission, this report summarizes where the POU evaluation efforts are meeting the basic rigor for EM&V studies in California and where the reports fall short. The data collected here forms the foundation for recommendations to improve the reliability of reported program savings, and ensure that consistent methods are being used by POUs to report program impacts.

---

<sup>8</sup> Ex ante is latin for “beforehand.” In models where there is uncertainty that is resolved during the course of events, the ex ante values are those that are calculated in advance of the resolution of certainty (e.g. third party measurement and verification of installed program measures).

<sup>9</sup> For the 2006-2008 IOU evaluations, a number of “high impact measures” were prioritized for enhanced rigor, based on the percent contribution to the overall portfolio.



## **Report Organization**

The remaining chapters of this report are organized as follows:

- Chapter 2 summarizes the EM&V reports in context of POU portfolio.
- Chapter 3 examines the EM&V reports across key components of a quality evaluation.
- Chapter 4 compares POU and IOU EM&V requirements.
- Chapter 5 concludes the report and provides recommendations.

## CHAPTER 2: Contextual Reporting Adequacy

The evaluation of POU energy efficiency programs is important to provide reliable estimates of program effects for Energy Commission energy consumption forecasts. As such, the evaluations should provide enough contextual information to provide the Energy Commission with an assessment of the reliability of the SB 1037 annual report claimed savings and any applicable adjustments to these claimed savings due to the evaluation efforts.

This chapter examines the evaluation reports submitted by POUs as independent assessments of measured savings, per AB 2021. To date, 11 POUs have submitted 13 evaluation reports. One study – *Evaluation of SMUD’s Prescriptive Lighting Program* – was found to encompass a measure persistence study, market potential study, and process evaluation to identify alternative program designs. Since the report did not include any verification of program claimed savings, it is excluded from this analysis.

Table 1 provides an overview of the evaluation reports submitted to date, including when the report was completed and for which program years.

**Table 1: Summary of POU Evaluation Efforts Submitted**

POU Name	Evaluated Program Year	Report Completion Date
Alameda Municipal Power	FY 2007/08	6/8/2009
Lodi Electric Utility	FY 2007/08	12/2008*(Taken from Document Properties)
City of Lompoc	FY 2008	3/10/2009
City of Palo Alto Utilities	FY 2007/08	2/19/2009
Port of Oakland	FY 2007/08	2/2/2009
Redding Electric Utility	FY 2007/08	3/25/2009
Roseville Electric	FY 2007/08	2/27/2009
Silicon Valley Power	FY 2007/08	3/20/2009
SMUD – HVAC	CY 2006/2007 <sup>10</sup>	3/31/2008
SMUD – Prescriptive Lighting <sup>11</sup>	Persistence Study of FY2001/2002	11/1/2007
SMUD – Refrigerator	CY 2006	5/1/2007
Truckee-Donner (TDPUD)	CY 2008 <sup>12</sup>	2/20/2009

<sup>10</sup> The savings estimates for the study were reported on a per-unit basis and not extrapolated to the total participant population. Data was collected in 2007, and included installations from 2006 and the beginning of 2007.

<sup>11</sup> Report is excluded from subsequent analysis because it does not provide an independent assessment of POU measure savings.

<sup>12</sup> It was not clear from report whether this is calendar year 2008 or fiscal year 2008, but the March 15, 2009, SB 1037 annual report indicates that the TDPUD programs are calendar year.

Turlock Irrigation District	FY 2008	3/16/2009
-----------------------------	---------	-----------

Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

## Portfolio Level Reporting

The POUs report portfolio level resource savings in the SB 1037 annual reports for net demand savings in Kilowatts (kW), net peak (kW) savings, and net annual Kilowatt hour (kWh) savings. Net lifecycle kWh and net lifecycle reductions in green house gases (GHG) reductions are also included, along with cost summaries and the total resource cost (TRC) ratio for the portfolio. The data is also broken out by end use and measure category, for example, heating, ventilation, and air conditioning (HVAC).

The EM&V reports, on the other hand, are generally organized by program, not end-use category. For instance, an EM&V report may include the non-residential HVAC program, but not the small business direct install program, which may include some non-residential HVAC measures. This would lead to an incomplete evaluation of the non-residential HVAC end-use category, but a complete evaluation of the non-residential HVAC program and its associated unique program processes and marketing strategy.

The EM&V reports do not always clearly state the ex ante savings<sup>13</sup> being claimed by the POU. This is not entirely uncommon for programs operating in jurisdictions new to energy efficiency programs and evaluation. In California, however, there is a well established system where the IOUs track unit accomplishments such as the quantity installed and apply deemed savings<sup>14</sup> to estimate ex ante program and portfolio savings. Since the POUs have the EE Reporting Tool to help them apply deemed savings to their programs, the POUs should also be able to easily estimate their ex ante claimed savings.

The relationship between the EM&V studies and the SB 1037 annual reports does not seem clear. In most cases, the energy savings reported in the SB 1037 report appear to be the POUs ex ante claimed savings – that is, savings which have not been independently verified, given the timing of the evaluations. One POU, Truckee-Donner, reported its evaluated (ex post) portfolio savings in the 2009 SB 1037 annual report for calendar year ending December 31, 2008.

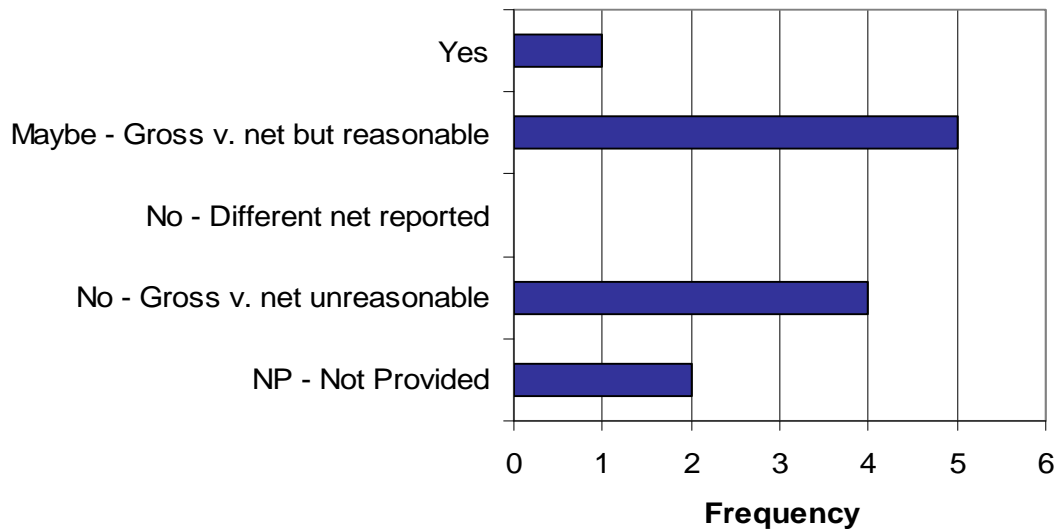
**Criterion 1.1** of the POU EM&V Framework examines the extent to which the EM&V reports provide sufficient context to understand how the programs being evaluated relate to the claimed savings being reported in the POUs SB 1037 annual reports.

Figure 1 summarizes whether the EM&V study reports (a) include any mention of the SB 1037 claimed savings, and (b) whether there is consistency between the reports.

---

<sup>14</sup> Deemed savings are the estimated energy savings that are expected to be achieved for different energy efficiency measures. The primary source of estimated energy savings is California's Database for Energy Efficiency Resources (DEER). <http://eega.cpuc.ca.gov/deer>

**Figure 1: Criterion 1.1 – Does the EM&V Study Report the Same Portfolio Savings Values as the SB 1037 Annual Report?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

The figure shows that only one EM&V report, Turlock Irrigation District (TID), included the same claimed portfolio savings (10,936,997 net annual kWh) as reported in the SB 1037 annual report.

Five of the 12 EM&V reports appeared to “maybe” have consistent savings, but it was not clear. The most significant source of confusion stems from gross versus net savings. The SB 1037 annual reports includes only the net savings, whereas the EM&V reports sometimes include only the ex ante gross savings and in other cases simply “kWh savings” with no indication whether the savings are net or gross. In one case, the evaluated ex post savings were found to match the SB 1037 annual report. Where POU's choose to report verified ex post savings in the SB 1037 annual report, it should be clearly indicated.

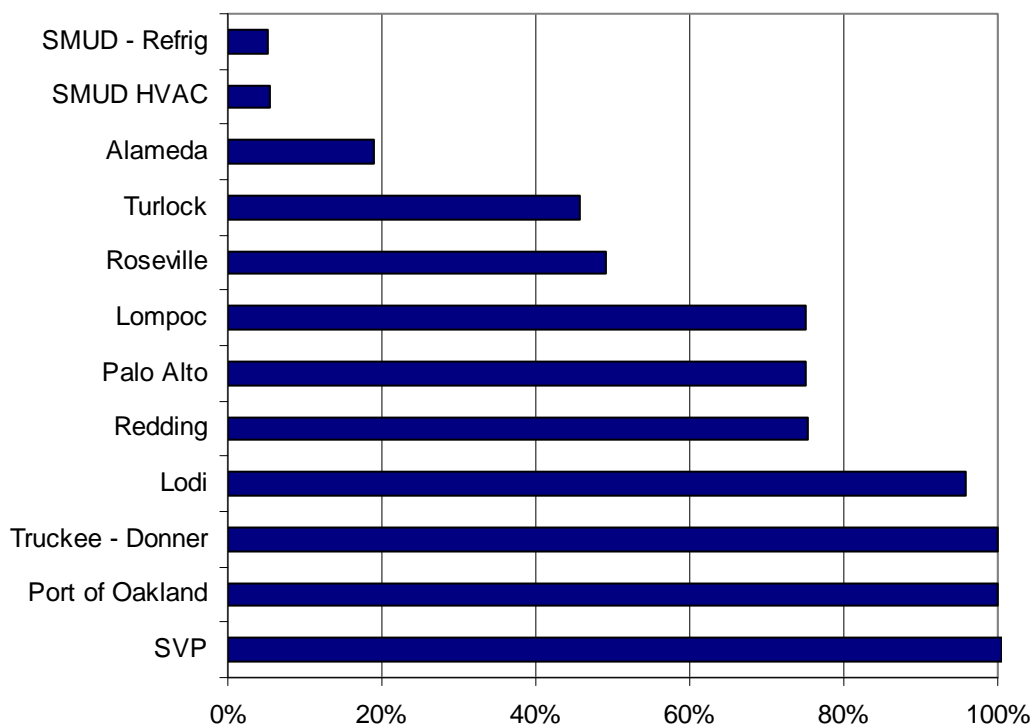
Four of the 12 EM&V reports provided ex ante savings, which appeared to be inconsistent with what was reported under SB 1037. For example, the Redding EM&V report included savings (not specified whether gross or net) for four measure categories. The total across the four measure categories was not provided in the EM&V report, but the summation came to 3,213,742 kWh annual, compared with the 1,639,577 net kWh reported in SB 1037.

Evaluation reports should clearly state the ex ante savings values (net and gross) associated with the specific programs being evaluated. The reports should also show how the program savings compare with the total POU portfolio savings, as shown in the SB 1037 annual report.

## Comprehensiveness of EM&V Efforts

In general, with the exception of SMUD, all POUs appeared to seek evaluation services to cover the majority of their portfolio of energy efficiency programs. **Figure 2** below compares the estimated percentage of SB 1037 savings included in each EM&V report.

**Figure 2: Criterion 1.2 – Approximate Percentage of Claimed SB 1037 Savings Covered by the POU Evaluations.**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Overall, reports for both SMUD programs, Alameda, Turlock and Roseville were vague regarding their findings relative to the entire portfolio.

- The savings from two SMUD programs that were evaluated, as calculated from the 2006 SB 1037, represent a combined 11 percent of portfolio savings. The SMUD reports were aimed at specific elements, and there was no intention that these studies were to be considered to be portfolio evaluations. No information was provided on why these specific programs were selected for evaluation, and what programs were excluded and their contributions to overall claimed savings.

- Alameda also evaluated a very low percentage (19 percent) of its portfolio because one large project that accounted for a majority of the portfolio savings did not allow any verification or evaluation activities. Otherwise, Alameda did evaluate the program that included almost all of the savings for the portfolio. Therefore, without this one problem site in the population, the evaluation would have covered a majority of the portfolio savings. Additionally, the large site did not receive any monetary incentive, so its inclusion in the program and associated reported savings are open for debate.
- Turlock had evaluation efforts that concentrated on its largest program but covered less than half of the portfolio. The Turlock report claimed that “almost 5,000,000 kWh” in reported savings were from two sites. It is clear from the statement that savings from these sites represent a significant percentage of portfolio savings and would be valuable samples to have among the population. However, this approach did not address what percentage of the total savings the two sites comprised. Overall, the report is too broad and lacks the necessary data to extrapolate the evaluation’s findings to the program or portfolio level.
- Roseville completed an evaluation that covered three programs. It is not clear what portion of the overall portfolio was included in the three programs. Ex ante savings were provided for one program – Commercial Custom Impact Evaluation (4,556,296 kWh saved). No ex ante savings were provided for the other two programs included in the evaluation, so there is no way to estimate what portion of the overall portfolio was covered. Regardless, the evaluation also did not provide ex post savings values for any of the programs.

The remainder of the EM&V reports submitted covered at least 75 percent of the portfolio savings reported in the SB 1037 report. For Truckee Donner, while 100 percent of the portfolio was evaluated, most of the evaluation was verification. Port of Oakland easily evaluated 100 percent of portfolio savings as the portfolio consisted of one site.

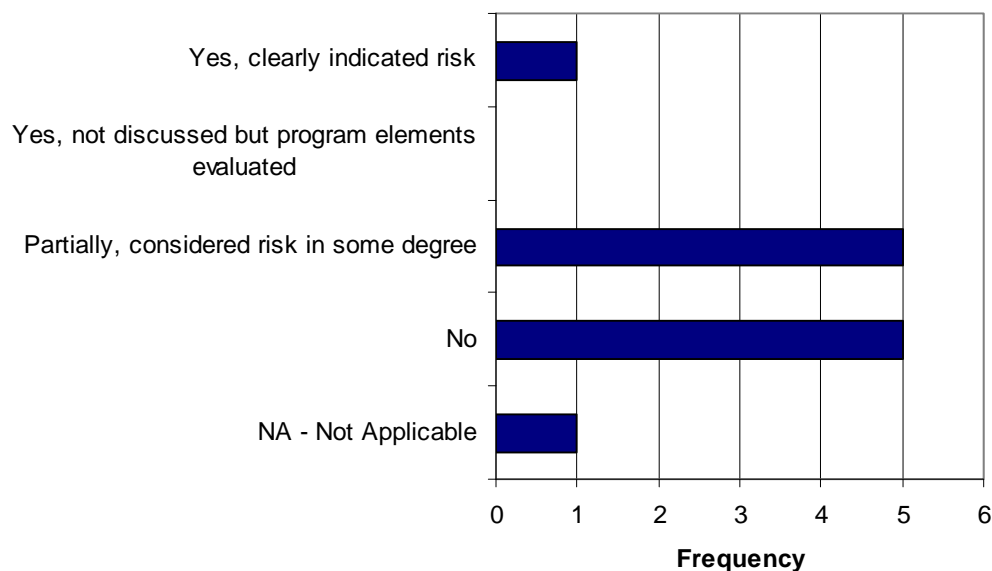
One final report, Silicon Valley Power (SVP), did not provide a percentage of total savings but was estimated to have captured the majority of savings. The report for SVP reported 24,663,638 kWh as the claimed savings from the non-residential program. Although the report does not compare this to the overall portfolio, a calculation using the 2006 SB 1037 of 24,509,440, would suggest it is more than 100 percent. While this may be comparing gross to net, the report does not approach the topic of net savings nor does it provide a percentage that is appropriately weighted to net or gross. Despite the lack of an accurate percentage, it appears that a high percentage of SVP’s portfolio was evaluated in this report.

Overall, most POU EM&V efforts appeared to commit the greater part of the study to the one or two programs that were responsible for the majority of portfolio savings. A supplemental process evaluation or scaled down impact evaluation was then used to evaluate a program or two of smaller savings. The remaining savings were either not mentioned at all, or excused from the evaluation by citing of a limited budget and scope of study.

It is expected that an impact evaluation evaluates and discusses a substantial percentage of the entire utility's portfolio. In some cases, a smaller portion of the total portfolio may be evaluated (or evaluated at a different level of rigor) if the remaining portion of reported savings are from well-established programs with robust evaluations in the same or similar utility service territories. The EM&V report should explicitly state the rationale for excluding portions of the portfolio from the (site-specific) evaluation and document the results from other evaluations that were used to derive the evaluated savings for the POU portfolio. The net-to-gross ratios and gross savings results may be most transferable. While installation verification rates are less transferable since they are more correlated to program implementation and quality control processes.<sup>15</sup>

For the most part the EM&V reports provided little discussion on the components of the POU portfolio that were not evaluated. **Criterion 1.3** was developed to determine whether the EM&V reports provided contextual information for prioritizing components of a portfolio for evaluation. **Figure 3** summarizes the frequency to which EM&V studies discussed the relative degrees of uncertainty and risk in the energy savings associated with different types of measures or programs.

**Figure 3: Criterion 1.3 – Discussion of Uncertainty Related to the Programs and Claimed savings Not Evaluated.**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

As displayed above, one POU (Palo Alto) addressed issues related to the uncertainty surrounding claimed savings not evaluated. Without specifically terming it “risk” or

<sup>15</sup> For example, IOUs may conduct site inspections of a random sample of projects as part of implementing their programs. If the POUs do not, then the verification results may differ.

“uncertainty,” Palo Alto’s EM&V report targeted a more complete collection of sites and purposely included the more customized and variable sites within the sample population. Because the report for Palo Alto included a discussion regarding its decisions to study specific programs and sites, appropriate steps were taken to minimize the uncertainty of the findings although the words “uncertainty” and “risk” were never used.

Five of twelve reports did not address risk or uncertainty in any context, while five reports addressed it implicitly meaning the evaluation contractor sought to include a large majority of the claimed savings for evaluation. Lodi, Lompoc, Redding, SVP, and Truckee all evaluated substantial portions of their respective portfolios. Despite the lack of a committed discussion of risk, the large inclusion of portfolio savings within the evaluation reduces much of the risk inherent in the report.

Ideally, a report would include a discussion that clearly identified the uncertainties of its findings, why they exist, and any attempts that were made to reduce them. If a report does not provide such a discussion, but clearly reduces uncertainty, then this is less ideal, but the risk is at least implicitly woven within the report’s methods. Such a method could target the programs of largest savings within a portfolio, could verify and survey the most variable and customized sites within a population, and could perform additional methods to ensure the riskiest reported savings are evaluated. Throughout the reports for Alameda, Roseville, both SMUD programs, and Turlock, not only does risk go unmentioned, but the methodology provided does not appear to evaluate the majority of the utility’s portfolio savings. These sites received a score of one because there is neither an implicit nor explicit consideration of risks or uncertainties.

In the Port of Oakland’s report uncertainty was not be applicable since the one site included in the portfolio was verified.

Throughout all reports, the topic of uncertainty did not receive sufficient discussion. Despite the attempt within several reports to implicitly minimize risk by evaluating 75 percent or more of the portfolio savings, a comprehensive discussion of risk and uncertainty would help to further contextualize the results.



## CHAPTER 3:

# Assessment of Measurement and Evaluation Studies

This section of the report examines the details of the EM&V reports for how well they meet the requirements of the *2004 California Evaluation Framework* and *2006 Evaluators' Protocols*. The primary purpose of the California protocols is to establish a uniform approach for:

- Conducting robust and cost-efficient energy efficiency evaluation studies.
- Documenting ex post evaluation-confirmed energy efficiency program and portfolio effects.
- Supporting the performance bases for judging energy efficiency program and portfolio effects.
- Providing data to support energy efficiency program and portfolio cost-effectiveness assessments.

Evaluation protocols may have other uses such as providing support for improving ex ante energy and demand estimates. Since the protocols were designed by CPUC to support the need for ratepayer accountability and oversight, the need for program improvements (especially cost-effectiveness improvements), and the documentation of effects from publicly funded energy efficiency programs in California, they are found to be applicable to the POU evaluations of energy efficiency programs.

## EM&V Studies Submitted and Reviewed

To date, 11 POUs have submitted EM&V studies per AB 2021. Twelve studies were found to include impact evaluations that provided independent assessment of measure savings. These 12 studies are briefly summarized below in Table 2. A full description of each EM&V report is included in Appendix A, including scores associated with each criterion.

**Table 2. Summary of EM&V Studies Submitted**

POU EM&V Report	Evaluated Program Year	Programs Evaluated for Impacts	General Approach
Alameda Municipal Power	FY 2007/08	Commercial Custom Program	5 largest rebated projects on-site verification and site specific savings re-calculation
Lodi Electric Utility	FY 2007/08	Non-residential Custom Program	All 5 lighting projects on-site verification and site specific savings re-calculation
City of Lompoc	FY 2008	Refrigerator Rebate Program Refrigerator BuyBack Program Income Qualifying Refrigerator Purchase Program	Sample of 21 replaced residential refrigerators. Paper verification and compared deemed savings to ENERGY STAR website deemed savings
City of Palo	FY 2007/08	Residential Refrigerator/Freezer	Paper verification (review of database)

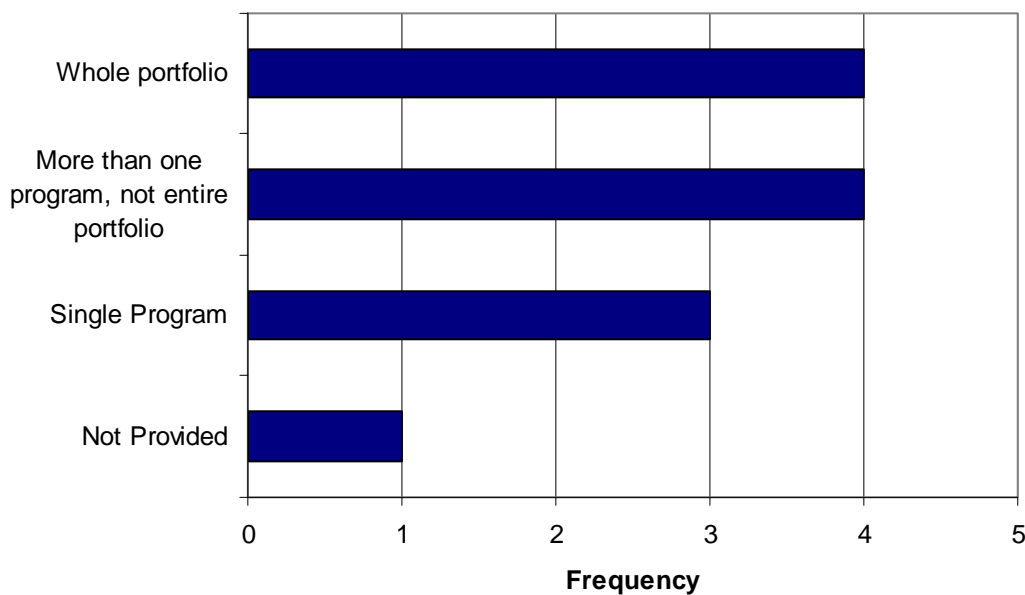
POU EM&V Report	Evaluated Program Year	Programs Evaluated for Impacts	General Approach
Alto Utilities		Recycling	and review of deemed savings
		Residential CFL Program	Sample of 50 participants for telephone verification; and review of deemed savings
		Right Lights Program	20 sampled projects on-site verification and site specific savings re-calculation
		Non-residential Custom Program	6 sampled projects on-site verification and site specific savings re-calculation
Port of Oakland	FY 2007/08	Non-residential Custom Program	1 site (census) on-site verification and site specific savings re-calculation
Redding Electric Utility	FY 2007/08	EarthAdvantage Program	70 sampled sites for paper verification. 1,252 sampled sites (census?) billing analysis for impacts by measure category
Roseville Electric	FY 2007/08	Residential New Construction	57 applications (out of 315) - Paper verification and discussion of relative merits of deemed savings estimates
		Residential HVAC Retrofit	57 applications (out of 350) Paper verification and discussion of relative merits of deemed savings estimates
		Commercial Custom Program	21 applications (census) Paper verification and discussion of relative merits of deemed savings estimates
Silicon Valley Power	FY 2007/08	Non-residential programs	10 sampled projects (out of 147) on-site verification, spot measurements, 1-2 week metering period
SMUD	CY 2006	Refrigerator Recycling Program	Participant telephone survey and in-situ monitoring
	CY 2006/2007	Residential HVAC Program	60 sampled participating homes – No verification, but metering to compare participant to non-participant energy use
Truckee Donner (TDPUD)	CY 2008	All 17 programs evaluated	Telephone survey and on-site verification, monitoring
Turlock Irrigation District	FY 2008	Non-residential rebate program	2 sampled sites on-site verification, 1 week monitoring, and site specific savings re-calculation

Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

## Scope and Rigor of EM&V Reports Submitted

As shown in the previous table, scope and rigor varied widely across the POU evaluation reports as did the extent of the “portfolios” and evaluation budgets. This stems from the diversity of California’s POUs, which vary greatly in size (both population and geographically), staffing, and energy efficiency experience. For some POUs, the implementation of energy efficiency programs itself is quite new, and they are still coming up to speed on evaluation requirements to independently verify savings.

**Figure 4: Criterion 2.1 – What is the overall approach of the EM&V report. Program by Program Approach, or Evaluation of Entire Portfolio (Measure Category or Customer Sector Basis)?**

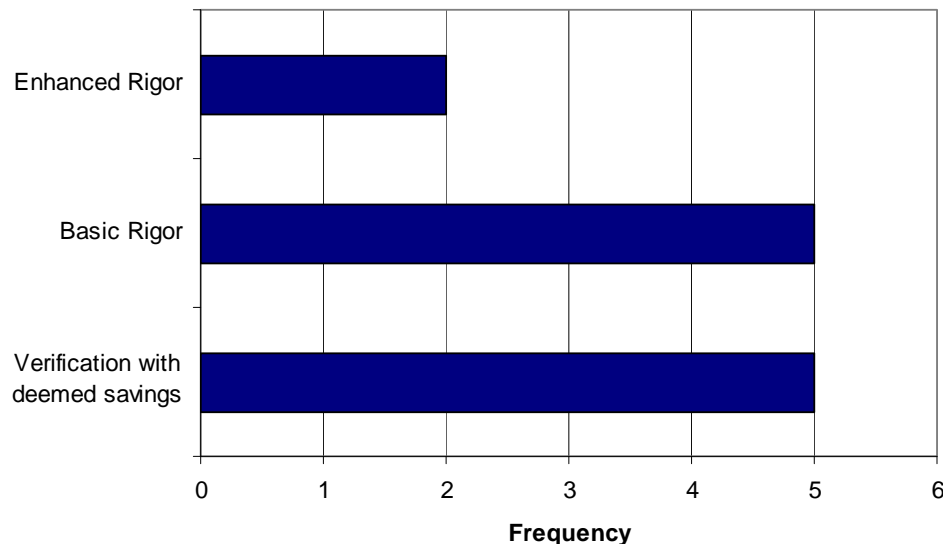


Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

For the most part, the POU evaluations generally assess either the entire portfolio of programs, or most of the portfolio. Although a complete portfolio evaluation can be one way to describe overall scope, the POU efforts varied widely. The range of portfolios that were evaluated ranged from a portfolio of one single project to the evaluation of 14 identifiable programs or distinct components. Alternatively, evaluations that focused upon single programs may have been appropriate if a single program accounted for the majority of savings to such a degree that other programs were inconsequential.

In addition to scope, **Criterion 2.2** examines the level of rigor the evaluations achieved and summarizes the overall approach of the EM&V efforts. **Figure 5** below shows the distribution of rigor level. Approximately half of the reports completed a paper or telephone verification with a review of the deemed savings values, and the remainder achieved at least a basic level of rigor.

**Figure 5: Criterion 2.2 – At What Level of Rigor Was the Evaluation Generally Conducted?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

**Figure 5** shows the highest level of rigor for the evaluation of any program or program element in the evaluation of the portfolios. That is, if a specific evaluation used verification, basic rigor, and enhanced rigor in the evaluation program elements, then it was scored as having achieved an enhanced rigor level. The figure above is meant to provide a high level assessment of general rigor achieved.

The two studies that achieved an enhanced rigor level were Palo Alto and the SMUD Residential HVAC EM&V reports.

- The Palo Alto report evaluated four programs, and used the *International Performance Measurement and Verification Protocols (IPMVP)* Option B for a project involving variable speed drives, which were installed on 12 exhaust and six supply fans for chemical fume hoods. Spot measurements were taken on as many units as possible (some power feeds were too difficult to access) and power meters were installed on five motors of varying sizes for a five-week period. This onsite and metering data was used to complete engineering calculations sufficient to meet IPMVP Option B criteria.
- The SMUD Residential HVAC evaluation was enhanced rigor because data collected included building characteristics, refrigerant charge test, system airflow, power testing, load monitoring, and temperature logging for participating site. Non-participating sites were measured for infiltration, total duct leakage, and duct leakage to the outside.<sup>16</sup>

<sup>16</sup> Total leakage includes leakage both inside and outside the building envelope. Leakage to the inside is a problem in that conditioned air is not being delivered to the proper area of the home, but it

Five studies achieved a basic level of rigor for at least one of the sampled sites included in the evaluation. In contrast to enhanced rigor, the basic level of rigor measures at least one engineering parameter, with assumed values for the remaining parameters.

- Alameda completed a billing analysis and verification with deemed savings review and self-reported hours. The billing analysis was completed at the basic rigor level, using whole facility analysis of utility meter data for the baseline and reporting period.
- Redding completed a billing analysis that appeared to comply with IPVMP Option C, which is whole facility analysis of utility meter data for the baseline and reporting period.
- The SMUD Refrigerator Recycling evaluation report used in-house monitoring of refrigerators and freezers that would be recycled.
- Silicon Valley Power evaluation took spot measurements of lighting circuits while toggling the time clock to determine which circuit powered a representative number of scheduled lighting fixtures. The evaluation also spot metered HVAC measure sites. For one site the evaluation used HOBO 4-channel loggers to log for one week, and for two more sites, the evaluation monitored HVAC units with current trend loggers for one to two weeks.
- Turlock Irrigation District also used an IPMVP Option A approach to its impact evaluation, with short-term metering used for the air compressor project site.

The evaluations were also heavy on installation verification with a review and critique of the deemed savings method primarily used for residential appliance rebates and non-residential lighting retrofits. For appliance rebates this approach is appropriate. Given the size of these programs and considering that numerous evaluations of appliance programs have been conducted elsewhere, it would not be cost-effective to duplicate efforts with a protocol-guided evaluation. However, non-residential lighting projects are a different story. The incremental cost of adding time-of-use logging and/or spot measurement to on-site verification is small enough that it is easily justified by the added confidence in the estimate. This is especially true when these projects account for a large percentage of the POU savings portfolios and that most of these programs have not been evaluated previously.

---

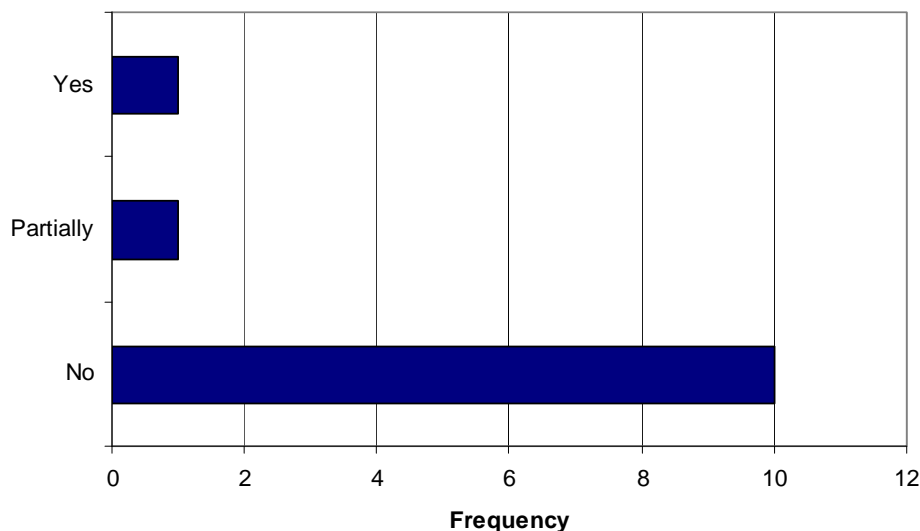
is delivered to conditioned space. Leakage to the outside includes only leakage *outside* the envelope, which increases system energy use since conditioned air that is leaked outside does not contribute to the heating or cooling of a home.

## Measure Life and Lifecycle Savings

Effective useful life (EUL) is defined as an estimate of the median number of years that the program's measures are expected to be operable. Lifecycle savings represent an estimation of the effect of programs over future years. **Criterion 2.3** assesses whether the EM&V studies account for lifecycle savings and whether any adjustments are made to EUL based on evaluation findings, mostly associated with operating hours.

A critical characteristic of measure savings is the estimate of how long the measure will persist after installation, commonly referred to as the EUL. Annual savings are given in units of kWh/year or therm/yr. Without an estimate of EUL, the real value of the measure is unknown, and cost-effectiveness can not be calculated. A comprehensive evaluation report should always address EUL at a minimum and as best practice should provide a table that shows the savings estimates over a given period of years. A table of that type was only found in one report, as shown below in **Figure 6**.

**Figure 6: Criterion 2.3 – Does the Evaluation Include an Assessment of EUL and Lifecycle Savings?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Only one EM&V report, Truckee Donner, adjusted EUL based on the evaluation findings – the ex ante EUL was 6.72 and the ex post EUL was estimated to be 7.27, assuming 8,000 lifecycle operational hours. The EUL was adjusted based on different annual hours of operation. Consequently, the lifecycle savings associated with the installed measure can be estimated for only one of the evaluation reports.

One other report partially addressed measure life, but in a very incomplete manner. Palo Alto's on-site verification revealed detailed instances of screw-in CFLs being removed prematurely, and stated that this could "severely affect the lifetime of these measures."

While the removal of CFLs is arguably more relevant to installation rate, a verification issue, partial credit is given here for mentioning measure life in this context.

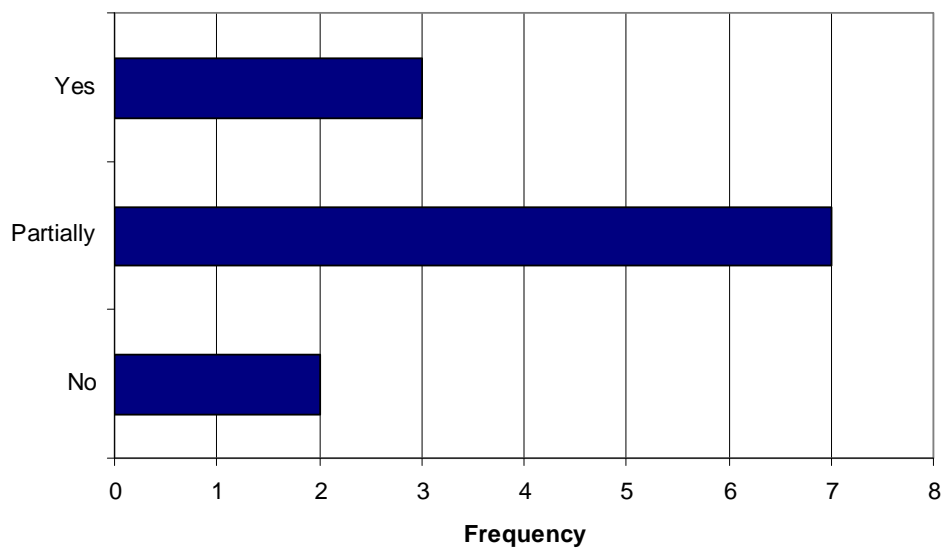
Overall, however, in nearly all of the POU evaluation reports (10 out of 12), there was no discussion or attention given to how long the installed measures were expected to last. The evaluation reports included only the annual kWh and kW impacts, with no reference to lifecycle energy impacts that were reported in the SB 1037 annual reports.

KEMA recommends that a lifecycle savings table be a prescriptive requirement of the evaluation studies. The table should extend through the end of the longest measure life in the program beyond the end of the evaluated program cycle. This will require some sort of estimate of effective useful life to be created or reviewed by the evaluator. Measure life for longer lasting equipment, such as chillers or boilers, is assumed to be approximately 20 years. Although the CPUC has provided the IOUs a standardized lifecycle savings table that extends to 20 years, there is now some discussion to account for longer measure lives, such as building shell measures, that can last more than 20 years.

## Documentation

For the most part, some documentation was provided in the EM&V reports to understand what was evaluated and how. **Figure 7** shows that only three EM&V reports were found to have sufficient documentation for a complete review.

**Figure 7: Criterion 2.4 – Does the Evaluation Report Provide Sufficient Documentation for a Complete Review?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

The EM&V reports with sufficient documentation were the SMUD HVAC, SMUD Refrigerator, and Silicon Valley Power evaluations.

- The SMUD HVAC report was a 153-page document that included equations for calculations, telephone survey instruments, and on-site data collection instruments.
- The SMUD Refrigerator report clearly details the data sources and methods used to estimate refrigerator and freezer energy use at different points in the appliance life cycle and under different operating conditions, data that were used to calculate ex post energy savings.
- The Silicon Valley Power evaluation included detailed description of the ex ante savings assumptions and site specific findings, including description of buildings, for example age, equipment size, operating hours, and parameters metered.

Seven studies partially provided sufficient documentation. These EM&V reports included Alameda, Lodi, Palo Alto, Port of Oakland, Roseville, Redding, and Truckee Donner. The documentation was insufficient because there were no engineering equations provided to understand how the evaluated savings were derived. Furthermore, there were no comparisons of parameter values from project application compared with the on-site verification results. For example, evaluation reports should provide tables that show ex ante fixture assumptions with actual fixture counts, comparison of wattage assumptions, operating hours, and kW consumption. Roseville provided a particularly thin evaluation, but the documentation was partially adequate because there did not seem to be much to explain; since no project specific impact analysis was completed, only a simple discussion of the ex ante savings methodology. Redding used a billing analysis to evaluate gross measure savings for four of the key program measures. While Redding provided some statistics from their regression models (realization rate coefficient estimates and their associated t-statistics), it did not provide complete model output, which would be helpful in assessing the overall validity of the analysis. Also, while the Redding analysis has the components necessary to provide estimates of total ex post savings, it did not include these estimates in their report, which would have been appropriate.

Two reports lacked sufficient documentation to fully understand what was done:

- The Lompoc evaluation did not include any discussion of the algorithm for estimating savings. Although the report explained the steps, there was no description of the overarching methodology
- For the Turlock evaluation, one of the two projects evaluated was an industrial food products manufacturing facility that upgraded its compressed air system and controls. Although the system was monitored for one week, the data was not provided and should have been for proper documentation.

In a robust evaluation report, the documentation should include all of the necessary details, such as assumptions and methods, such that the reader does not have to guess at the assumptions made and reverse-engineer calculations. A properly documented evaluation should be comprehensive enough that another reasonably competent consultant could recreate the analysis with another dataset. This does not mean the evaluator needs to share every line of code produced for the analysis; it means that the approach is well-defined and described clearly in the report and that all of the methodological steps and calculations are clearly stated.



Unfortunately, insufficient and partial documentation were characterized in most of the POU evaluations. While basic approach and some explanation were stated, the pathway from raw data to the final results was not clear in most cases. A considerable number of evaluations gave few clues beyond general site observations, with no clear explanation of parameters adjusted and how the ex post savings were calculated.

## **Gross Savings Approaches Utilized**

There are two types of savings estimates that are normally desired from impact evaluation: gross savings and net savings. Gross savings are calculated for *all* the technologies and measures installed for program participants and included in the program tracking database. Net savings control for savings that would have occurred for these participants over the same period regardless of whether the program was offered.

The criteria discussed in this section focus on the proper estimation of gross savings. The “Impact Evaluation Protocol” provided in the *2006 California Evaluator’s Protocol* is applicable to programs claiming energy or demand savings and for programs that are expected to influence energy-related behavior. The “Impact Evaluation Protocol” was developed to ensure that all evaluations of program-specific impacts are conducted using evaluation methods. The protocol also guides the estimation of evaluation-adjusted gross and net savings for energy (kWh) and demand (kW) for electricity-using equipment.

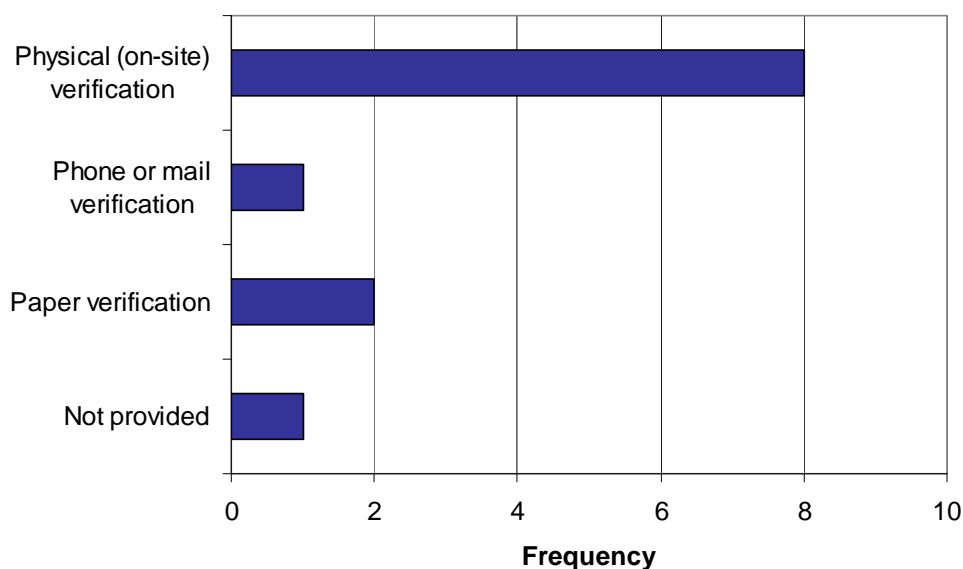
One of the studies reviewed, the SMUD Residential HVAC evaluation, did not calculate gross savings. It included on-site data collection at both participant and non-participant homes and calculated net savings directly, rather than calculating gross savings and applying a net-to-gross ratio. For that evaluation, the gross savings criteria were more appropriate for evaluating the net savings analysis than the net-to-gross criteria discussed in the following section. Therefore, the net savings approach for the SMUD Residential HVAC evaluation was evaluated based on the criteria in this section and is presented in the following tables with the results for the gross savings approach of the other studies.

Some of the studies reported results for evaluations of multiple programs, often with different approaches and levels of rigor. The ratings provided below for each criterion represent the highest level achieved in each report. For example, if a report evaluated three programs in different ways, and one included paper verification, one included phone verification, and one included on-site verification, Criterion 3.1 (verification) would be rated a 3 (on-site verification), based on the highest level used in the study. Detailed information on each evaluation is reported in Appendix A.

## Verification

The *California Evaluation Framework* generally requires that measure installation verification be included as a component of the overall measurement and verification approach. The objectives of measure installation verification are to confirm that: (1) the measures were actually installed, (2) the installation meets reasonable quality standards, and (3) the measures are operating correctly and have the potential to generate the predicted savings. Figure 8 summarizes the types of verification used by the studies.

**Figure 8: Criterion 3.1 – Does the EM&V Report Include Suitable Measure Installation Verification?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

A majority of studies included physical verification of installed measures (on-sites). One study, the SMUD refrigerator recycling program evaluation, did not report any verification activities.<sup>17</sup> The remaining studies either performed a paper verification, consisting of a review of program tracking data, applications, and invoices to ensure consistency and measure eligibility, or phone or mail verification, which ask participants about measure installation and retention.

As noted previously, the results of the on-site verification were not clearly documented to show any differences in fixture counts or other parameters such as operating hours. Additionally, documentation of the on-site data collection forms and telephone survey instruments should be included in the EM&V report.

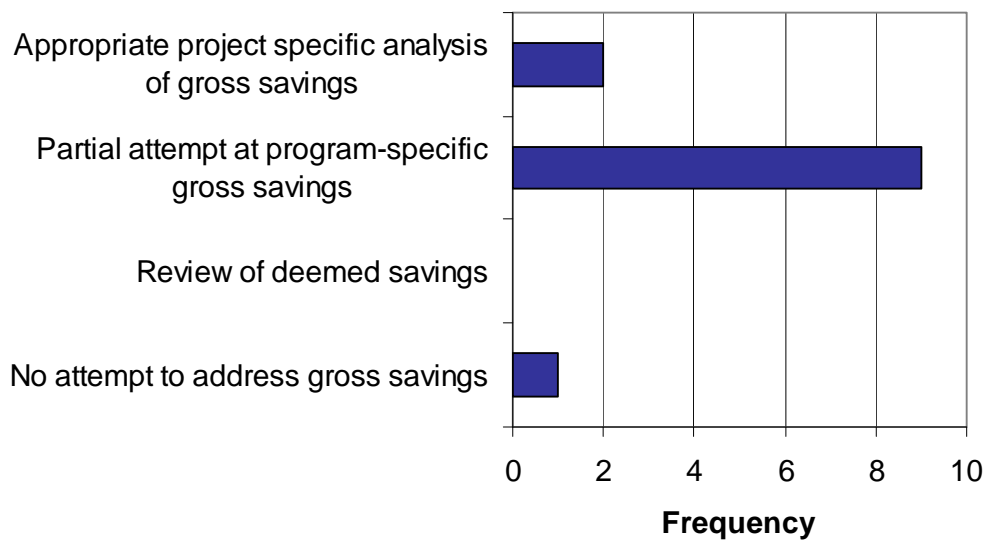
---

<sup>17</sup> Refrigeration recycling and verification activities would determine if the refrigerators and freezers claimed by the program were, in fact, picked up from the customer and recycled, not trashed or resold.

## Gross Savings Methods Used

The studies reviewed employed a range of approaches and rigor levels. Most studies were done at the basic level of rigor according to the International Performance Measurement and Verification Protocols (IPVMP). The most commonly used approach, as defined in the IPVMP, was “Option A,” consisting of field measurement of key parameters, with additional parameters being estimated. Figure 9 represents an assessment of the suitability of the gross savings evaluation method. All but three of the studies included an estimate of program-specific gross savings.

**Figure 9: Criterion 3.2 – What Gross Savings Evaluation Method was Used?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Only two studies were found to have performed an appropriate analysis of gross savings.

- The SMUD HVAC collected detailed on-site measurements at a sample of participant and non-participant homes, and net savings were estimated directly by comparing the results from the two groups.
- The SMUD Refrigerator Recycling program used the results in situ metering done for this evaluation with in-situ metering results from and IOU evaluation and data on at-manufacture energy use to create an accurate estimate of energy savings for the retired appliances.

The majority (9 out of 12) studies were found to have partially attempted program specific gross savings, mostly by conducting field verification and on-site inspection of basic project parameters, for example, the operating hours based on occupant self-report. In most cases, no information is provided related to on-site sampling approaches, or whether all units were verified through the on-site visit.

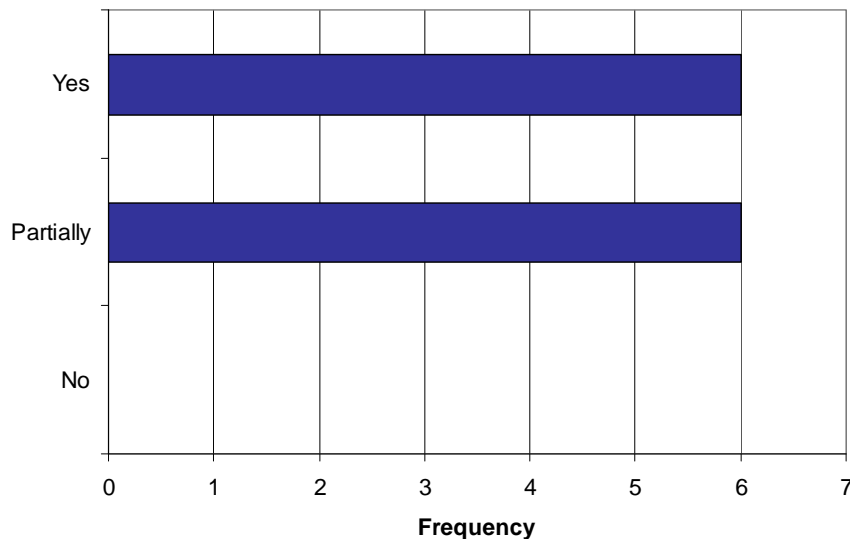
- Alameda also completed an on-site verification with site-specific adjustments to annual operating hours based on facility personnel interviews.
- The Lodi study conducted on-sites activities at five lighting sites, using information on daily and seasonal schedules to adjust simple engineering models.
- The Lompoc study collected energy use data for a sample of specific refrigerators recycled by the program from secondary sources and compared them to deemed savings.
- The Port of Oakland study conducted an on-site evaluation, resulting in an adjusted measure count and adjustments to measure wattage assumptions.
- The Redding study used a billing analysis to estimate savings impacts of residential HVAC and shell measures.
- The Turlock Irrigation District's study used an engineering analysis to evaluate a lighting project and short-term metering to evaluate a compressed air project.
- The Palo Alto, the Silicon Valley Power, and the Truckee-Donner studies all evaluated a variety of measures using different approaches, including reviewing deemed savings, simple engineering models, billing analysis, and short-term metering or monitoring.

Finally, Roseville included a review of deemed savings but no site verification. There was some debate as to whether this constituted an impact evaluation at all, or was simply a review of ex ante savings. Of all the studies, its approach was the most inadequate for evaluating ex post savings because the report discussed the ex ante savings methods and provided no adjustments based on program reported achievements.

## Baselines

Since gross electricity savings are based on a comparison of energy use associated with the measure installation relative to some baseline pattern of use, the selection of baseline condition is of utmost importance. The selection of an appropriate baseline is usually contingent on the type of technology being installed and the intent of the program. For example, normal replacement baseline is appropriate when the unit being replaced was at the end of its useful life and would have been replaced in the same approximate time frame in the absence of the program. Early replacement baseline is appropriate when the unit being replaced would have remained in operation in the absence of the program. If early replacement is used, then remaining useful life of baseline equipment should be considered and explicitly stated. **Figure 10** summarizes the evaluation of the EM&V reports discussion and choice of baseline.

**Figure 10: Criterion 3.3. Is the Baseline Suitable?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Half the studies were found to have suitable selection of baseline for savings calculations. Lodi, Redding, Roseville, and the two SMUD studies were all found to have used a suitable baseline, while Palo Alto was found to have used a suitable baseline for one of the components of the evaluation. The two SMUD studies both included analysis and comparison of multiple baselines, with a clear explanation of which baseline was used for the impact estimates and why. While the choice of baseline was found to be suitable for Lodi, Redding, and Palo Alto they provided at best only cursory discussion of the issue.

The remaining reports were found to have partially suitable baselines, mostly due to little or no discussion of the issue. However, the EE Reporting Tool provides applicable baseline assumptions (natural or early) for specific measures. If the POUs are selecting the correct measures associated with their program installations, then the ex ante baseline should be

generally correct. The evaluators should review the baseline assumptions associated with the deemed savings and interview program participants during the verification visits to verify baseline assumptions (for example, fixture wattage previously installed and remaining useful life of equipment that was replaced before its end of life).

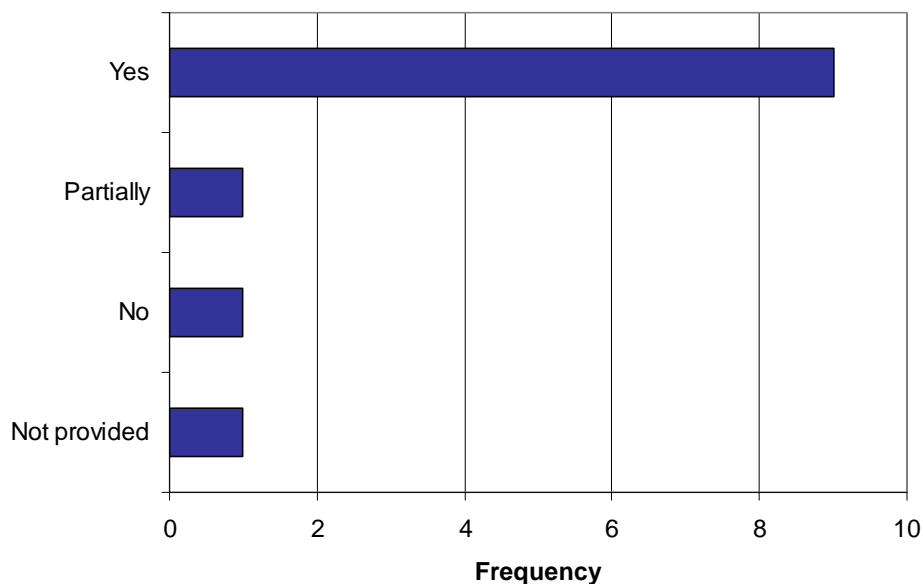
## Sampling

All evaluation studies collect data from participants, non-participants, or the market to provide information for evaluation analysis. Unless all relevant members of a group have data collected from them (a census), some type of sampling is used in order to cost-effectively complete evaluations.

The goal of the sampling and research design of an impact evaluation is a sound, defensible, unbiased determination of the actual gross and net savings for the overall program. Some measurement error is acceptable for each sample project – especially if the measurement error is small relative to the sampling variability. But measurement bias should be minimized since it will propagate through the analysis.

The following graph presents the assessment of the sampling approaches used in the EM&V studies.

**Figure 11: Criterion 3.4. Is the Sampling Approach Appropriate?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Most of the studies (9 out of 12) used an appropriate sampling approach, including 2 that included all participants (census). Lodi and Port of Oakland completed a census of their program projects. The remaining seven evaluation reports generally considered statistical significant in selecting sample size. For instance, the Roseville evaluation included a verification of 16 randomly selected projects, which represented the statistical confidence of

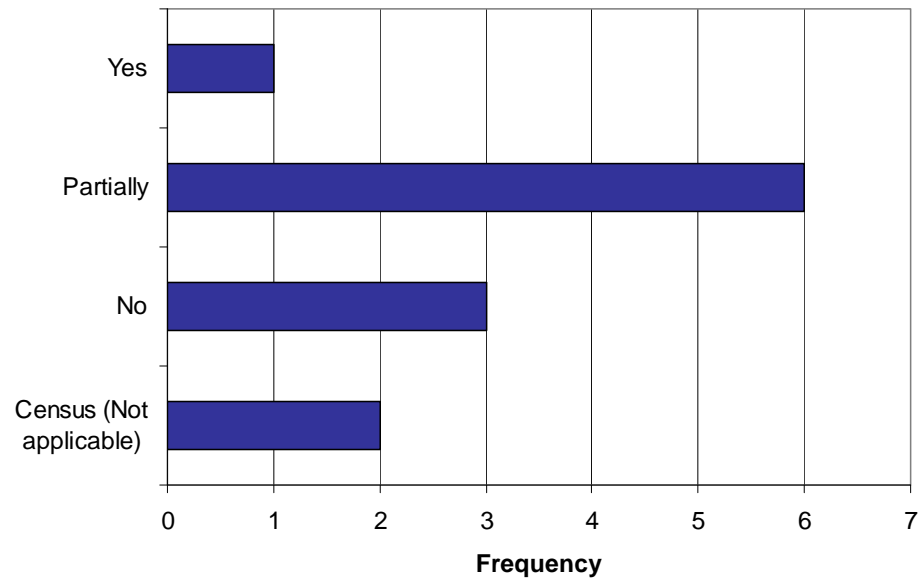
90 percent+/-10 percent. Redding also completed a paper verification of 70 installation, representing the same statistical confidence, along with a billing analysis of 1,252 customers that was deemed to be statistically significant.

The remaining EM&V reports used sampling approaches that had some issues:

- Alameda was given a “partially appropriate” score, having surveyed four of eight participants. The evaluation had to omit the largest project in the program given lack of customer cooperation, but the project was so large that without it the sample could not be considered appropriate.
- Turlock was given a “not appropriate,” since two sites were evaluated, but no information was provided about the population. The two sites could have represented a census or only a small fraction of the participants. While the sites chosen were large and represent a significant share of both the non-residential and overall portfolio, it would be inappropriate to extrapolate from the results to the overall population.
- The final study, the SMUD Refrigerator evaluation, was rated a Not Provided (NP), as it pooled new metered data with data collected in other studies, but did not provide enough information about either sample to make an assessment.

While the sampling approach for most of the studies seemed appropriate, few of the studies presented sampling precision targets. Such targets – and the corresponding achieved precision – are necessary to assess the reliability of the resulting savings estimates are. A specified confidence interval and desired relative precision should be used to determine the size of the sample necessary to meet those targets. Where precision is calculated from multiple evaluation study efforts, the precision information should be provided for each study effort as well as the combined result.

**Figure 12: Criterion 3.5 – Were the Sampling Precision Targets Stated and the Achieved Precision of the Estimates Reported?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

The above graph shows that only one of the studies (SMUD HVAC evaluation report) reported targeted and achieved relative precisions appropriately.



Six studies partially addressed precision, typically by reporting precision targets, but not presenting the achieved precision of the completed study. In one study, both targeted and achieved precision were reported, but the achieved precision was implausibly low, 0.008 at 90 percent confidence (meaning the estimate was extremely precise), and insufficient detail was provided to verify the calculation. For the studies that reported a target precision, they ranged from 80/20 to 90/10. Table 3 shows the targeted and achieved precision for each study.

Three studies (Alameda, SMUD Refrigerator, and Turlock) contained no discussion of the precision of the estimates, although each of the EM&V report used a sampling approach of some kind.

- Alameda sampled four of the largest projects, with no information on total population size.
- SMUD Refrigerator reported the number of households (and associated refrigerators and freezers) sampled for free-ridership questions for net-to-gross, with no mention of confidence level or precision or information related to total program population.
- Similarly, Turlock sampled 2 sites that represented nearly 5,000,000 kWh (or 50 percent of total claimed non-residential energy savings).

Two studies (Lodi and Port of Oakland) included all participating sites (census), so a discussion of precision was unnecessary.

**Table 3. Target and Achieved Precision by Study**

<b>POU EM&amp;V Report</b>	<b>Target Precision</b>	<b>Achieved Precision</b>
Alameda	NP	NP
Lodi	NA (Census)	NA (Census)
Lompoc		
- By program	80/20	NP
-All programs combined	90/15	NP
Palo Alto	90/10	NP
Port of Oakland	NA (Census)	NA (Census)
Redding	NP	NP
Roseville	90/10	NP
Silicon Valley Power	80/20	NP
SMUD HVAC	90/12.6	90/12.9
SMUD Refrigerator Recycling	NP	NP
Truckee Donner	90/10	0.0008
Turlock Irrigation	NP	NP

## Execution

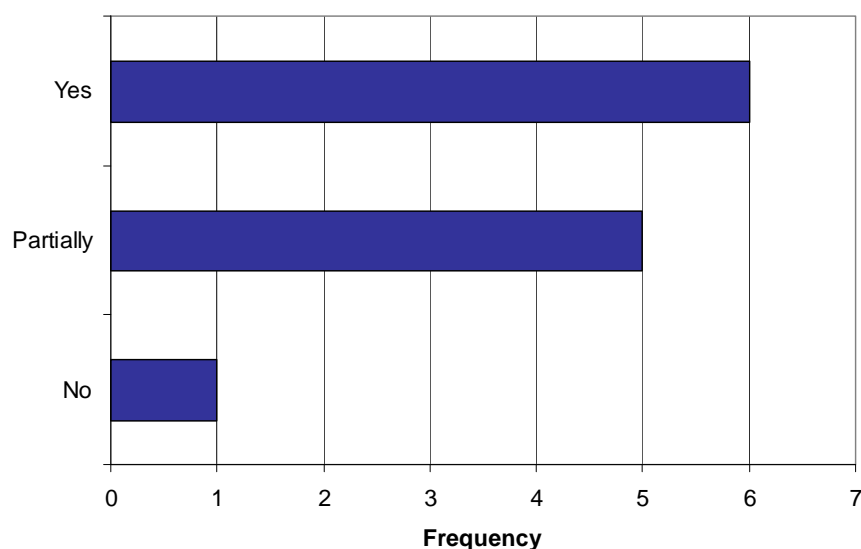
Each EM&V study was assessed on the execution of the selected savings method. For deemed savings reviews the evaluation should reference appropriate sources, preferably specific to California (where possible). The Database for Energy Efficient Resources (DEER) is considered a robust source of deemed savings values.

For engineering analysis, the appropriate engineering algorithms should be used with key parameters specified. Were correct baseline and retrofit efficiencies used (where correct federal standards referenced)? If metering was used, did it occur during the appropriate season for weather-sensitive measures? The reviewer also checked that energy savings were properly annualized to a full year and that interactive effects between measures were accounted for.<sup>18</sup>

For billing analyses, reviewers checked that adequate data was used. Typically, 12 months of pre- and post-installation data is recommended, with a minimum of 9 months being used. Reviewers checked to see that the models include the appropriate variables and that variables representing program participation were included correctly. If a control group was used, the reviewer assessed whether it seemed like a reasonable control group.

The following table shows the breakdown of how well the savings approach was executed.

**Figure 13: Criterion 3.6. Was the Selected Savings Approach Executed Appropriately?**



<sup>18</sup> For instance, the analysis of energy savings must include consideration of heat/cool interaction in conditioned spaces. This is particularly important for lighting measures in conditioned spaces and other measures in conditioned spaces that reduce internal heat gains.

Half (six) of the studies properly executed their selected approach: Lodi, Palo Alto, Port of Oakland, Redding, SMUD HVAC, and SMUD Refrigerator. For the most part, these studies incorporated site visits to confirm installation rates and interview program participants on key equipment operating parameters, for example, operating hours, fixture wattages, and temperature set points. Redding completed a billing analysis that first segmented the customers that were likely to have a statistically significant result from the billing analysis. The SMUD Refrigerator evaluation included a participant survey to examine what would have happened to the refrigerator or freezer in the absence of the program. This is an important component to examining refrigerator and freezer recycling programs.

The EM&V studies that received partial credit include Alameda, Lompoc, Silicon Valley Power, Truckee Donner, and Turlock Irrigation District. The reasons for the partial rating are explained below.

- The Alameda billing analysis used a very coarse average daily temperature approach by month rather than a more precise and accepted cooling degree day approach. Furthermore, the results were extrapolated from just a few months of post-implementation data that produced an unrealistic trend line. For Site 2, the evaluation contractor found that while the calculated savings were more accurate (based on the site's actual operating hours), the deemed savings resulted in a higher savings number and subsequently the report recommended that the site claim the deemed savings value. This is not an acceptable verification of project savings.
- Lompoc also completed a refrigerator recycling program evaluation but suffered from a flawed assumption that overstated gross savings and ignored field findings that would have shown considerably less savings. Specifically, the evaluation assumed that all recycled refrigerators would have been used as a second refrigerator. This is not an appropriate assumption, as other scenarios include being kept but not used, discarded to landfill, and discarded to the secondary market where use may be less than the entire year.
- Silicon Valley Power appeared to have the right approach, but rather than developing the more accurate estimates of energy consumption, the report passively identified but did not resolve the erroneous assumptions held within deemed savings such as incorrect operating hours.
- The Truckee-Donner evaluation had assessments of 14 portfolio components that were evaluated with varying degrees of rigor. Most of the well-documented elements appear to have been executed properly, but considerable gaps in documentation for other components made it difficult to determine whether the verification and evaluation results was properly calculated. Additionally, the relative precision statistics of 0.008 for the study are suspiciously low.

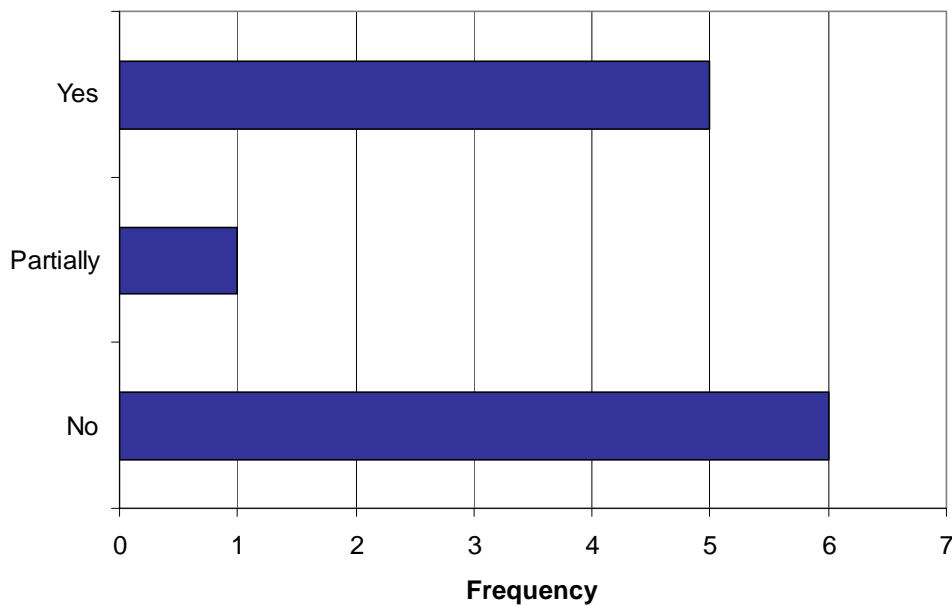
- The Turlock study's evaluation of the compressed air site raised more questions than it resolved. The evaluation identified significant seasonal variation in energy use and used interviews with facility staff to seasonally adjust metered data from this study and an earlier metering effort. However, the presentation of the adjustment calculations is unclear, and the documentation for the assumptions is incomplete. In particular, the usage measurements cited in the evaluation are noted as having been taken "on the cusp of the transition in [seasonal] use," but in the explanation of the adjustments, these values appear to have been taken to represent typical high-season usage, an assumption that appears to be unsupported.

Finally, the Roseville EM&V report included a review of deemed savings, but this was found not to have been executed properly. The Roseville EM&V report reviews the deemed savings approach used by the POU but provides "findings" such as "we recommend analyzing the updated DEER savings numbers when they are released." In another case, the report reviews the building simulation approach and provides a recommendation that the program account for savings related to surpassing Title 24 code by 20 percent. No revised savings estimates are provided for any specific measures, much less for projects that participated in the program. The EM&V report should have conducted a few site visits to verify installations and attempted to re-calculate the deemed savings based on the site specific findings. At the very least, the evaluation report should have provided some revised deemed savings numbers based on improved values from secondary literature sources. No such results were provided.

## Expansion of Sample to the Population

The goal of sampling and research design is to estimate actual gross savings *for the overall program* in a sound, defensible, and unbiased way. Evaluation studies use sampling approaches to focus limited evaluation budgets cost-effectively to improve the overall quality of the study. If a proper sample design was used, then the results of the sampled projects can be used to develop estimates of program population results that will be close to the true values that would have resulted if the same data collection had been carried out for all projects.

**Figure 14: Criterion 3.7 – Are the Results Extrapolated to the Program Population, in an Adequate Way?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Although five EM&V reports are scored as having expanded the results of the sampled sites back to the program population, two evaluations, by default, were applicable to the program population because the sites represented a census.

- Port of Oakland, with its one site, was included as having its sampled results applicable to the program as a whole.
- Lodi was included as having extrapolated the five lighting sites results to program wide, but the five lighting sites were a census for that measure category, and there were other measure types included in the Non-residential Custom Rebate Program (although project count and savings were not provided in the report).

Furthermore, it was found that Palo Alto expanded the sampled results for a subset of programs that were evaluated for only the Smart Energy Program and Right Lights

program. Otherwise, the following two evaluation reports were found to have completed the expansion appropriately:

- The evaluation for Lompoc completed a deemed savings review and adjusted the estimated savings based on the model type and serial number information and obtained energy use from the database of refrigerators maintained by the ENERGY STAR ® program. The revised savings values were applied to all program participants to derive a total verified savings value for the program.
- The SMUD Refrigerator evaluation also expanded the results back to the population of program participants.

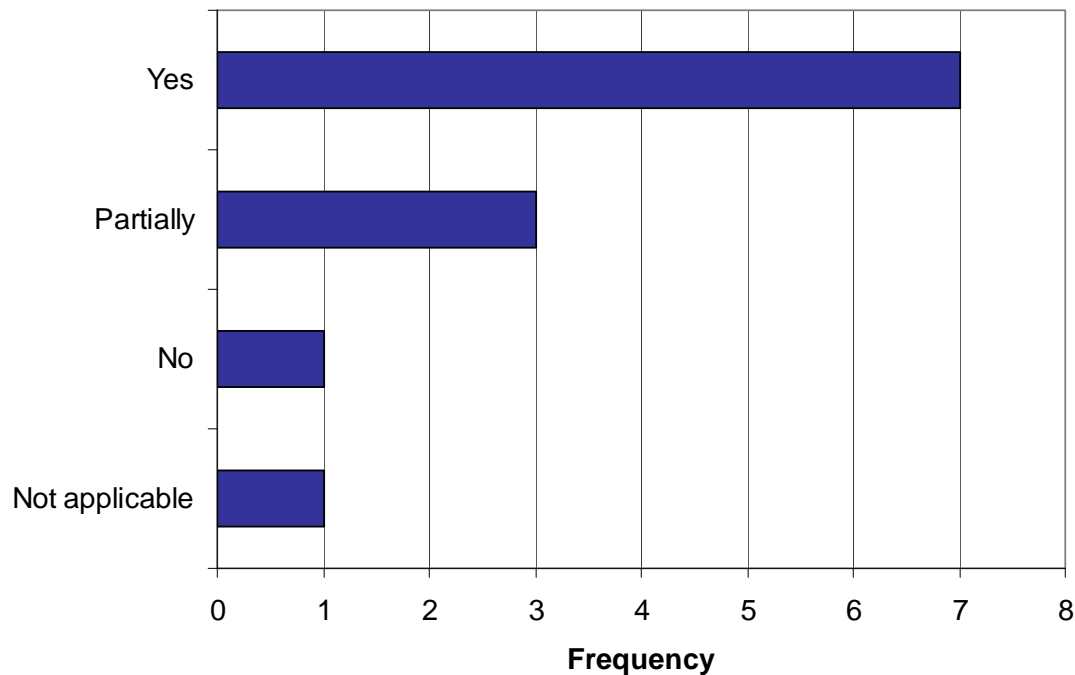
One study, Truckee-Donner, was given partial credit for insufficient documentation to indicate the expansion method. It appeared to have used a simple random sampling where the unweighted realization rate of the sample is applied to the population, but that can't be determined from the report. Typically the weighted realization rate should be applied.

Unfortunately, half of the EM&V studies did not make any attempt to expand the evaluation results of the sample to the program population. These studies provided realization rates only for the sampled sites, for example , Site 1, Site 2, and Site 3, but did not provide any meaningful results at the program level. At a minimum, the program savings should be adjusted to reflect the verified savings from the impact evaluation, with an explicit note that the remainder of program savings were passed through from the ex ante claim.

## Discussion

The final criterion on which the EM&V studies were evaluated was whether the report provided adequate discussion and comparison of the ex ante and ex post savings estimates. Since the purpose of evaluation is to provide meaningful feedback on ex ante assumptions, this is an important component to improving program design and measure mix for future years.

**Figure 15: Criterion 3.8 – Does the Report Explain Differences Between Ex Ante and Ex Post Savings Estimates?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Seven of the twelve studies provided adequate explanation. These studies included Lodi, Palo Alto, Port of Oakland, SMUD HVAC, SMUD Refrigerator, Truckee-Donner, and Turlock. Examples of reasons for discrepancies include revised wattages from the program tracking data based on field findings, revised operating hours for equipment, and differences in quantity found to be installed.

The following specific examples from the Port of Oakland are provided to illustrate reasons for the verified savings to differ from the ex ante savings:

- *“Inconsistencies between the reported fixtures on the application and the actual installation. The itemized invoices provided with the applications showed adjustments from the reported installation. However, some of these changes may have been made during installation.”*

- *“Incorrect reporting of fixture wattages.* The wattages for both the four and six lamp T5 high output fixtures and some of the 400 watt metal halides units were misreported. Standard wattages for these fixtures are available and could be provided to vendors to fix this problem.”

Three of the studies provided some discussion, but not enough for the reader to completely understand the difference in savings.

- Alameda provided some discussion of discrepancies in narrative (paragraph) format: however the discussion is hard to follow, and savings table would make it great deal easier to understand.
- Silicon Valley Power identified limitations to specific site findings but claimed incorporating those limitations into an ex post calculation was beyond the scope of the study. In one site with three 20 horse power (hp) motors and one 30 hp motor, the assumed annual operating hours of a 20 hp motor were different than those of the 30 hp motor. The reality was that all four pumps were operating on the same schedule. The limited scope of the project was described as the reason that this discrepancy would not be used in the calculation of ex post savings. The study’s scope was acknowledged as the limiting factor that prevented the development of more accurate ex post savings in four of the ten case studies. Two other studies were unable to draw complete conclusions because anticipated data was not able to be located by the time the study was published.
- Lompoc evaluation study provided some discussion of the differences between their verification findings and the ex ante values, but then suggested that Lompoc use “E3 replacement values” as reasonable although the E3 values generated a program estimate 33 percent higher than the evaluator’s field findings.

One contained no discussion – Redding completed a billing analysis but did not provide any context for how the billing analysis results should be interpreted relative to the original deemed savings estimate. Although the billing analysis results were close to the ex ante estimate a brief discussion of the deemed savings methodology may be helpful to provide insight on why the values are different. Reasons for the difference may also be related to actual cooling degree days relative to the typical meteorological year that is used for building simulations of weather sensitive measure savings.

One study was considered to be not applicable, since the report did not provide any improved values, this criterion is deemed not applicable to this EM&V report. Roseville’s evaluation report included a short discussion for areas where the deemed savings may be improved, such as updated DEER or PG&E workpaper, but no revised values.



## Net-to-Gross

The overall goal of impact evaluations is the determination of energy and demand savings induced by the program. In other words, the savings need to be “net” of what would have occurred in the absence of the program. Hence EM&V studies need to identify free ridership, that is, what participants (and non-participants) would have done in the absence of the program that would affect their energy use level. While an apparently simple concept, in over two decades of experience in the evaluation field, it has proven to be difficult to objectify free ridership. The table below summarizes the California protocols related to the determination of participant net impact evaluation.

**Table 4. 2006 Evaluators’ Protocols Required Protocols for Participant Net Impact Evaluation**

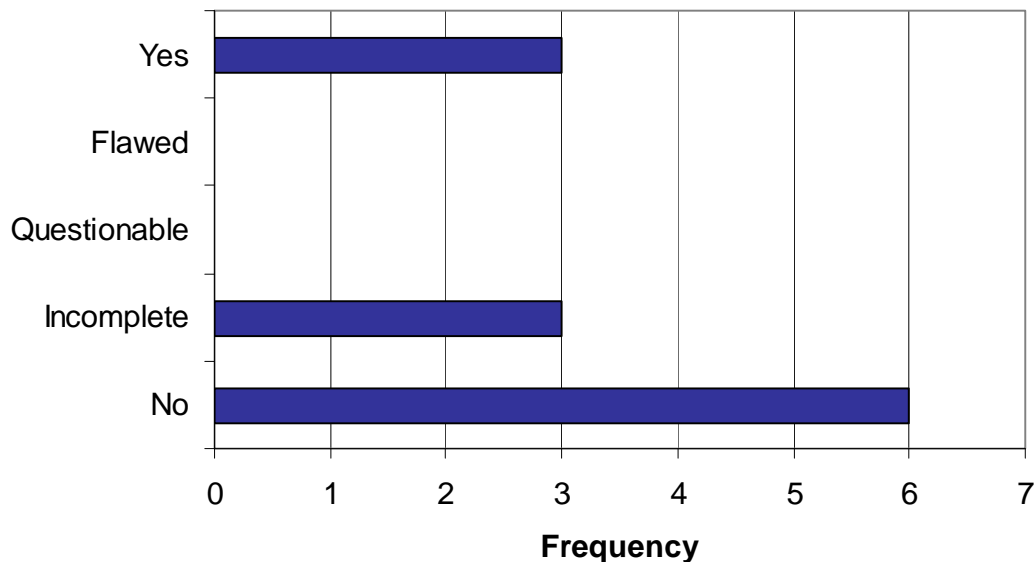
Rigor Level	Minimum Allowable Methods for Participant Net Impact Evaluation
Basic	1. Participant self-report.
Standard	<ol style="list-style-type: none"> <li>1. Participant and non-participant analysis of utility consumption data that addresses the issue of self-selection.</li> <li>2. Enhanced self-report method using other data sources relevant to the decision to install/adopt. These could include, for example, record/business policy and paper review, examination of other similar decisions, interviews with multiple actors at end-user, interviews with mid-stream and upstream market actors, Title 24 review of typically built buildings by builders and/or stocking practices.</li> <li>3. Econometric or discrete choice<sup>30</sup> with participant and non-participant comparison addressing the issue of self-selection.</li> </ol>
Enhanced	1. “Triangulation” using more than one of the methods in the Standard Rigor Level. This must include analysis and justification for the method for deriving the triangulation estimate from the estimates obtained.

Source: California Energy Efficiency Evaluation Protocols: *Technical, Methodological, and Reporting Requirements for Evaluation Professionals (“Evaluators’ Protocol”)*. Prepared for the California Public Utilities Commission. April 2006. [http://www.calmac.org/events/EvaluatorsProtocols\\_Final\\_AdoptedviaRuling\\_06-19-2006.pdf](http://www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf)

Being in compliance with the basic rigor defined in the Participant Net Protocol requires the estimation of a net to gross ratio (NTGR), based on participant self-report at a minimum. The participant self-report generally requires the development of a survey instrument, scoring for responses and handling of missing data and inconsistent responses to develop the NTGR.

As can be seen in the graph below, fully 6 of the 12 studies did not discuss NTG in any way.

**Figure 16: Criterion 4.1 – Does the Evaluation Include a Quantitative Assessment of Net-to-Gross?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Of the six studies that addressed net-to-gross quantitatively, three were found to adequately document net savings (Truckee Donner and both SMUD evaluations).

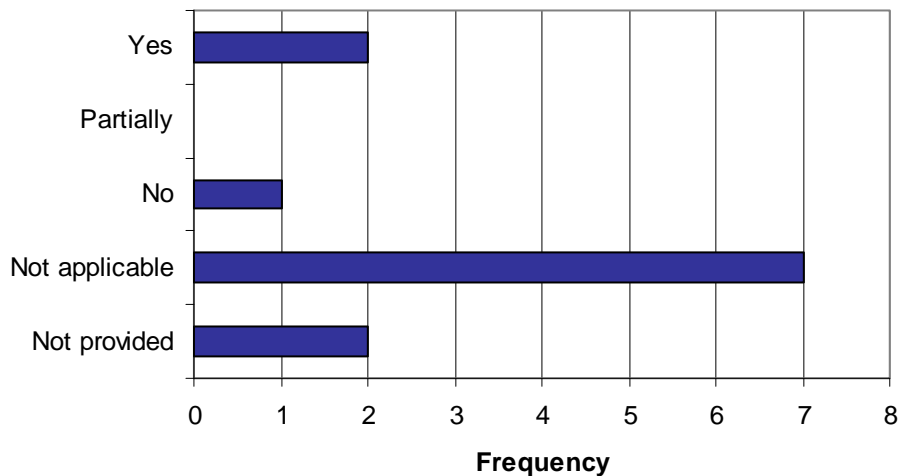
- The SMUD Residential HVAC evaluation included on-site data collection at both participant and non-participant homes and calculated net savings directly, rather than calculating gross savings and applying a net-to-gross ratio. This is considered an enhanced approach to net-to-gross determination. The study also used a conventional participant survey to estimate a free ridership for the residential HVAC program. This estimate of free ridership was not used to estimate net savings (or to estimate gross savings), but was provided for SMUD to help them understand the market for their program. The criteria in this section were evaluated based on the design of the free ridership survey.
- The SMUD Refrigerator Recycling evaluation used a self-report survey of program participants to estimate free-ridership. The questions were related to what the participant would have done with the refrigerator or freezer in the absence of the program.
- The Truckee Donner study also reviewed the applicability of deemed NTGR from the EE Reporting Tool using secondary data sources. A deemed NTG value is appropriate if the programs are similar in customer base and if program delivery strategy is comparable.

The other three studies that addressed net-to-gross quantitatively did not develop net savings estimates. Two studies (Alameda and Palo Alto) used self-report surveys to explore free-ridership and offered recommendations based on the findings but failed to develop a net-to-gross ratio (NTGR) or net savings. The study for Redding offered program design recommendations based on the most recent DEER values for the NTGR of evaluated program measures, but failed to apply the NTGR to gross savings to develop net savings.

### Sampling and Precision

Evaluations typically seek to achieve an expected statistical precision of  $\pm 10$  percent or  $\pm 20$  percent at the 90 percent confidence level. In many cases, it is generally appropriate for the NTG sample to be the same as the gross impact sample, if the sample is appropriate for the gross impact analysis. **Figure 17** below summarizes the assessment of the sampling approach used by the EM&V studies. For the reports that did no NTG analysis, this criterion is considered not applicable.

**Figure 17: Criterion 4.2. Is the Sampling Approach Appropriate?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Both SMUD studies used a sufficient sample size to determine the net-to-gross ratio within these bounds.

- For the SMUD HVAC study, the free-ridership survey covering both equipment and Title 24 code compliance aspects was completed by 60 participants, either as part of an on-site or by telephone.
- SMUD's refrigerator recycling study used telephone surveys of 203 households and stratified its sample into five categories to determine the NTGR.

Alameda's evaluation was deemed not to have employed an appropriate sampling approach, since it asked four out of the total five large commercial participants about free-

ridership. However, the evaluation report chose not use the survey to generate site or program net-to-gross ratios, citing insufficient data.

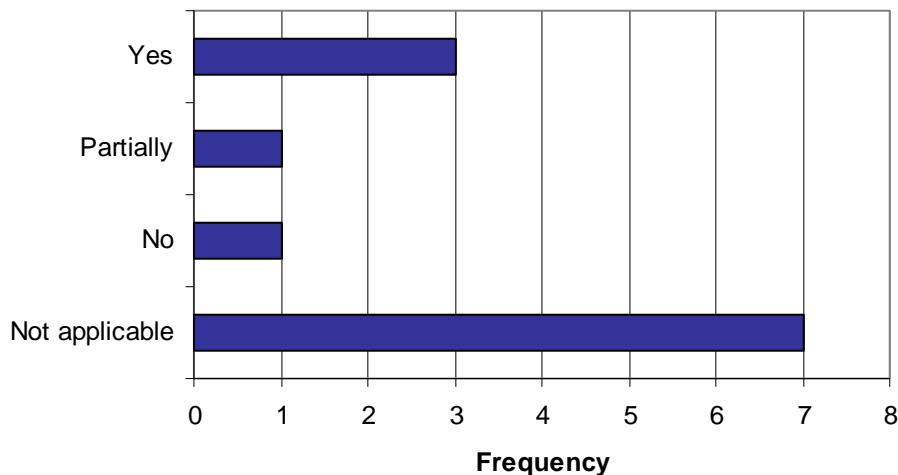
Two evaluations did not provide sufficient information related to the net-to-gross sampling approach. Truckee-Donner and Palo Alto evaluations both used telephone surveys as well but did not document their sample sizes sufficiently to be evaluated.

The Redding evaluation employed a literature survey to review deemed NTG values and thus fell into the “not applicable” category with the six remaining studies.

### Net-to-Gross Methodology

The survey method is the most straight-forward method of free-ridership estimation, and the lowest cost method. It does, however, have its disadvantages in potential bias and with accuracy. A well-constructed survey attempts to minimize bias by employing questions from a variety of different perspectives to “triangulate” the most accurate answers, and employ a consistency check question to adjust the individual’s estimate accordingly. In order to handle “don’t know” answers, missing data, and inconsistent answers, the *California Evaluation Framework* recommends dropping respondents from the analysis.

**Figure 18: Criterion 4.3 – Is the Selected NTG Analysis Method Applied Appropriately?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Both SMUD studies and the Truckee-Donner evaluation deployed surveys consistent with the 2006 *California Evaluators' Protocols* in this regard.

- In the SMUD HVAC evaluation, net savings were calculated through a direct comparison of participant and non-participant savings based on logger data and field measurements. As such, the equipment and code compliance NTG ratios derived from the telephone and onsite surveys were not used to calculate net savings from gross savings. The study presents both NTG ratios, notes the difficulty in isolating bias in the code compliance section of the survey, and leaves the

application of the equipment NTGR up to SMUD's discretion. In cases like this, however, it is preferable for the evaluators to make a decision, since they are in the best position to judge the strength of the different approaches.

- The SMUD refrigerator recycling survey had three questions addressing free ridership, including whether the respondents would have kept or gotten rid of the appliance in the absence of the program, and what they would have done with the appliance (stored unplugged, kept it running, gotten rid of by recycling, traded in, or sold).
- Truckee Donner used a net-to-gross participant survey with key questions associated with timing of the program participation relative to the decision-making process. That is, the questions examined instances where participant awareness of the program caused them to purchase and install the efficient measures. Other questions included influence of the program, and self-reported actions that would have been taken in the absence of the program.

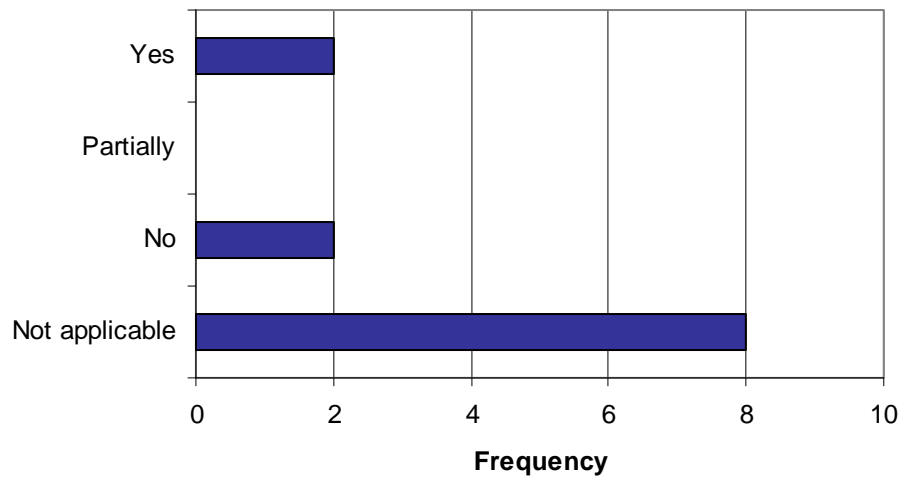
The Palo Alto EM&V study received partial credit for its NTG method. The survey posed two questions on free-ridership: (1) whether the participant had considered purchasing the efficient measure earlier, and (2) what the likelihood was (using a five-point scale) of purchasing the more efficient equipment absent the program. The number of respondents by technology was as follows: CFLs (50 customers); refrigerators (20 customers); appliances (24 customers); HVAC (2 customers); water heaters, pool pumps, and insulation (3 customers per technology). While the insights into CFL free-ridership were valuable for program managers, the small sample sizes and lack of questions examining free-ridership from more than one perspective meant the evaluation did not meet the Framework criteria.

The Alameda study did not apply the NTG analysis method appropriately, despite the fact that some free-ridership questions were asked of the four out of five large commercial participants during on-site visits. One of the four participants indicated he would have installed the equipment without the program, while two said the program was the major reason for the installation. One participant was not sure. The EM&V report made no adjustments to the gross savings estimate citing the "qualitative nature" of these findings. The specific questions were not included in the EM&V report as documentation. The net-to-gross estimation could have been improved by using a more rigorous net-to-gross battery of questions that included quantitative ratings, for example, a rating from 0 to 10 of how important the incentives were, whether the participant would have installed the same quantity or efficiency level.

Redding employed a literature survey to review deemed NTG values and thus fell into the "not applicable" category with the six remaining studies.

In using the self-report method, the NTG analysis should also account for partial free-ridership, which recognizes that participation in a program influences behavior in ways beyond simply whether to purchase a more efficient measure. The partial free-ridership issue is generally addressed through additional questions about stated intentions that are contingent on the response in the primary question. These questions are only asked of those that probably or definitely would have taken the actions.

**Figure 19: Criterion 4.4 – Does the Approach Account for Partial Free-Ridership?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

The Truckee-Donner and SMUD HVAC evaluations were the only EM&V reports to have captured partial free-ridership in a series of survey questions regarding the decisions governing the installation timing and efficiency level of the measures in question.

Two studies that quantified net-to-gross did not quantify partial free-ridership.

- The Palo Alto survey asked what the likelihood would have been, absent the program, of purchasing the efficient measures. The question used a five-point scale to gauge customer responses, which is an appropriate design. However, as the survey did not capture the programs effect on the quantity or timing of the purchase of the efficient measures, it was insufficient to properly account for partial free-ridership.
- The evaluation for Alameda also included a qualitative assessment of free-ridership. Since no discussion of partial free-ridership was included in the report this analysis presumes that partial free-ridership was not addressed.

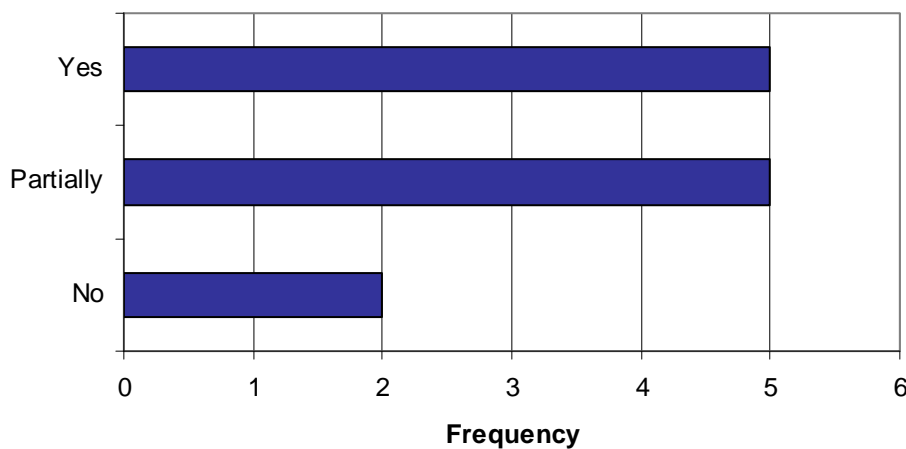
Partial free-ridership is not applicable to the SMUD Refrigerator Recycling program, as there is no standardized approach for these types of programs. Free-ridership is a complicated issue for refrigerator recycling programs, since the program influences the disposal of equipment, rather than the installation of new measures. The free-ridership issue is already captured in the survey related to behavior in the absence of the program, but methods to quantify whether the participant was influenced to act sooner is not well defined. Furthermore, questions related to disposal of more units are not applicable, as most households will dispose of only a single refrigerator or freezer at a time.

## EM&V Summary and Conclusions

Although the primary purpose of evaluation is to document the amount of ex post or net energy saved through the programs and to provide information to help determine the cost-effectiveness of acquiring those resources, the evaluations should also provide information to help improve programs and assist in making the best possible choices in a public policy context. Therefore, an important component to EM&V reports is to provide clear and actionable recommendations for how to improve program operations.

The EM&V reports should provide an understanding of why the observed effects occurred and identify ways to improve program effectiveness to achieve energy savings cost-effectively. Where discrepancies exist between the ex ante claimed savings and ex post evaluation savings estimates, the EM&V report should provide recommendations for improving realization rate. **Figure 20** below summarizes how many of the POU EM&V reports provide recommendations that would be useful for improving program processes.

**Figure 20: Criterion 5.1 – Does the EM&V Report Provide Clear Recommendations for Improving Program Process to Improve Realization Rates and/or Verified kWh Results?**



Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

Five of the EM&V studies were found to provide relatively robust recommendations for improving program processes. Recommendations may be categorized in several types:

- Program measures modifications – Additions or removal of specific measures. (For example, “consider providing a \$25 rebate for ENERGY STAR televisions and DVDs, and eliminate rebates for dishwashers.”)
- Program marketing and messaging – How to frame the program and program benefits (For example, “incorporate more non-energy benefits into program messaging, including home comfort, safety, and environmental benefits.”)

- Program documentation – What documentation should be provided to improve energy impact estimations and/or to facilitate verification of installed units. (For example, include hours of operation in the application form, or a specs sheet showing the efficiency of the installed air conditioner.)
- New programs – Additions to efficiency program portfolio (For example, consider adding a refrigerator recycling program) along with initial ideas for partnering opportunities such as the California Waste Removal Systems Buy-Back Center & Appliance Recycling Company.

In some cases, the evaluation contractor, Summit Blue, provided the same recommendations to each POU, including recommendations for adding a \$25 rebate for ENERGY STAR Digital-to-Analog Converter Boxes, televisions, and DVDs.

Where recommendations are provided, the EM&V report should also include an explanation and rationale for the recommendation. This may seem obvious, but it was not always found to be included in all EM&V studies.

**Criterion 5.2 “Does the evaluation provide an assessment of the reliability of the verified savings?”** was the final criterion developed for the Framework. The reported energy savings associated with POU efficiency programs will inform the estimated effects on Energy Commission energy consumption forecasts. As such, the assessment of the reliability of verified energy savings is important for resource planning. None of the EM&V reports were found to specifically address the reliability of verified savings beyond the statistical significance determinations.

The evaluations should include an overall judgment of the reliability of the verified savings, including areas of uncertainty that may have affected the evaluation results. Historically, evaluations have generally provided statistical precision estimates but have not rigorously addressed engineering and measurement precision. The California Public Utilities Commission (CPUC) is looking to include engineering and measurement precision for IOU evaluations, with efforts underway with the Master Evaluation Contractor Team (MECT) Engineering Working Group. CPUC requires site-specific EM&V reports to discuss general uncertainties that may affect evaluation results. For the most recent IOU evaluations for 2006-2008 program years, there were varying degrees of addressing the reliability of the evaluated savings, and even for the CPUC, this is still a new issue being explored.



## CHAPTER 4:

# Comparison of POU and IOU Evaluations

This chapter provides a discussion of recent IOU evaluation activity and how the POU evaluations relate to the IOU evaluations. The analysis considers evaluation budgets, realization rates and net-to-gross ratios, and evaluation reporting requirements.

### Budget Considerations

For the 2006-2008 IOU energy efficiency programs, the CPUC set EM&V budgets at about 8 percent of total program funding (\$163 million). For the 2010-2012 IOU programs, the budgets were reduced to 4 percent of total program funding (\$125 million). These CPUC budgets set reasonable guidelines for POU EM&V budgets, with the following qualifiers:

- IOU-related EM&V budgets include funding for overarching planning and policy support studies, and evaluation budgets that do not include these types of activities may be somewhat smaller.
- IOU evaluations achieve some economies of scale, for example, it usually takes less than twice the budget to evaluate a program that is twice as large, and therefore budgets for smaller programs may need to be somewhat proportionally larger.

Table 5 presents selected POU program and proposed evaluation budgets for the FY 2008 period. Program savings and budgets come from the 2009 SB 1037 Report, and EM&V budgets come from POU evaluation plans posted on the NCPA website, usually reflecting budget ranges. Although the table shows results for all POUs that have submitted evaluation plans, it is not known what the final evaluation budgets were. The actual evaluation budgets may have deviated substantially from the proposal submitted by the evaluation contractor. The Energy Commission may consider requesting a “Final Evaluation Plan” that includes the targeted sample sizes and actual methods to be used by the evaluation contractor, as well as actual evaluation budget. The evaluation proposals that were submitted may not be sufficient as an “Evaluation Plan.”

The last two columns of the table relate the proposed EM&V budgets with total program spending. The rows highlighted in green show POUs that have proposed EM&V budgets that fall below the lower 4 percent evaluation budget target that was set by CPUC for EM&V of IOU programs during the 2010-2012 period.

As the table shows, 5 of the proposed EM&V plans were highlighted as falling below the 4 percent EM&V budget target. Since these utilities are also associated with some of the larger program budgets, the overall EM&V spending for the 13 POUs that submitted proposed evaluation plans falls between 2 percent and 3 percent of total program spending.

Eight of the POUs filed proposed EM&V plans that have evaluation budgets at or about the 4 percent to 8 percent guidelines that have been used for the IOU programs. These POUs tend to have smaller programs, where evaluation budgets might need to be proportionately higher to provide minimum thresholds for an effective evaluation.

Overall, it appears that POU evaluation funding is on the low side and may need to increase somewhat if POU EM&V activities are to keep pace with IOU evaluation activities. Also, some POUs might consider combining EM&V activities to gain some economies of scale. This would be especially appropriate for POUs that run similar programs.

**Table 5. POU Energy Efficiency and EM&V Budgets**

POU Name	FY 2008 Reported Net kWh savings (SB 1037)	FY 2008 Reported Total efficiency program spending (SB 1037)	FY 2008 EM&V proposed budget range (from plans)		EM&V budget % of Total FY 2008 Program Spending	
			Low	High	Low	High
Biggs	132,877	\$40,027	\$3,000	\$5,000	7%	12%
Gridley	23,550	\$53,541	\$10,900	\$17,800	20%	33%
Healdsburg	236,349	\$119,884	\$10,900	\$17,800	9%	15 %
Lassen	123,046	\$147,889	\$12,000	\$19,000	8%	13%
Lodi*	3,090,527	\$414,649	\$31,000	\$47,000	7%	11%
Lompoc*	304,163	\$122,884	\$5,400	\$9,300	4%	8%
MID	1,870,992	\$437,549	\$50,000	\$60,000	11%	14%
PSREC	595,600	\$386,495	\$12,400	\$19,900	3%	5 %
Redding*	1,639,577	\$2,304,669	\$25,000	\$40,000	1%	2%
Roseville*	9,313,572	\$2,057,660	\$49,840	\$49,840	2%	2%
SVP*	24,509,440	\$5,803,153	\$56,646	\$56,646	1%	1%
TID*	10,936,997	\$1,144,259	\$25,000	\$40,000	2%	3%
Ukiah	278,721	\$105,440	\$10,900	\$17,800	10%	17%
Total	53,055,411	\$13,138,099	\$302,986	\$400,086	2%	3%

\* Indicates that a final evaluation report was submitted

Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

While this analysis would ideally be able to compare the quality of EM&V effort with the evaluation budget, this is not possible due to the lack of information on actual budget spent. Furthermore, not all POUs listed in the table had submitted final EM&V reports. For the POUs where final EM&V reports were submitted (indicated with an asterisk in the table), Silicon Valley Power (SVP) was found to have provided the most robust evaluation of programs with a review of the non-residential programs across four measure categories (lighting, HVAC, motors/VFDs and other). None of the EM&V reports for POUs included in the table, however, had completed a sufficient evaluation of program level impacts, meaning the verified savings were applicable only to the sites evaluated.

From the assessment of EM&V reports, the funding provided for the evaluations may have been insufficient for the evaluation contractor to fully address all components of the POU portfolios, including using at higher levels of rigor for more complex and uncertain projects, metering of equipment and system performance for longer periods, and completing net-to-

gross data collection and analysis. With the budgets provided, however, the evaluation contractor should have been able to provide contextual information related to the program(s) evaluated, document where the ex ante savings were passed through and where savings were adjusted, and provide an overall ex post portfolio savings result.

## Realization Rates and Net-to-Gross Ratios

Although the overall rigor and scope of the POU evaluations varied significantly, most EM&V reports provided a realization rate of some kind. A realization rate represents the ratio of the ex post evaluated savings to the claimed ex ante savings. Therefore, a 100 percent realization rate means that the ex post savings matched the ex ante savings perfectly. For the California IOUs, both gross and net realization rates are often provided. The gross realization rate compares gross evaluation findings (including both install rate and savings estimate) with ex ante gross savings estimates. The net realization rate compares net evaluation findings (gross evaluation savings multiplied by the net-to-gross ratio) with ex ante net savings estimates. Table 6 provides a summary of the realization rates that were documented in the EM&V reports and the scope associated with the realization rate. In most cases, the realization rate did not include net-to-gross ratio because net was not evaluated.

**Table 6. Summary of POU EM&V Realization Rates**

POU Name and Specific Program	kWh Realization Rate	kW Realization Rate	Applicability
Alameda – Commercial Custom	82%	98%	Weighted across 5 sampled sites, no NTG
Lodi – Non-residential Custom	92%	95%	Weighted across 5 sampled sites, no NTG
Lompoc – Refrigerator programs	n/a	n/a	Provides review of deemed savings, with no program project specific analysis
Palo Alto – Refrigerator	99%	n/p	Program, but no realization rate for demand
Palo Alto – CFL	100%	n/p	Program, but no realization rate for demand
Palo Alto – Right Lights	88%	71%	Weighted across 20 sites, no NTG
Palo Alto – Custom	131%	137%	Weighted across 6 sites, no NTG
Port of Oakland – Non-residential Custom	108%	108%	One site
Redding - EarthAdvantage	96%	n/p	Billing analysis, no realization rate for demand, no NTG
Roseville – 3 programs	n/p	n/p	Paper and phone verification with no verified savings results
Silicon Valley Power – Non-residential	101%	n/p	Weighted across 13 sites, no NTG, no realization rate for demand
SMUD – Refrigerator	n/p	n/p	Provides analysis of measure savings, but not applied to program
SMUD - HVAC Tier 1	117%	69%	Program level net only, comparison of ex ante net with ex post net
SMUD - HVAC Tier 2	120%	67%	Program level net only, comparison of ex ante net with ex post net
Truckee Donner – Portfolio	114%	255%	Across 17 programs
Turlock Irrigation District –	98%	98%	Weighted across 2 sites, no NTG

Non-residential			
-----------------	--	--	--

Source: CMUA's *Energy Efficiency in California's Public Power Section – A Status Report*, March 2010 and Northern California Power Agency (NCPA) <http://www.ncpa.com/energy-efficiency-m-v-reports-2.html>.

The CPUC recently completed evaluations for the California IOUs 2006-2008 programs. These evaluations covered an array of programs and measures. Table 7 displays some key findings of these evaluations, organized by evaluation contract group. Both the realization rate and the underlying net-to-gross ratio is provided. These results were mainly gleaned from a review and analysis of the evaluation reports' executive summaries. In several cases tables in the bodies of reports were also accessed.

**Table 7**  
**Key Parameters from CPUC Evaluation Reports for 2006-2008 IOU Programs**

Evaluation	Component	Realization Rates			Net-to-Gross Ratio			Notes
		kWh	kW	Therms	kWh	kW	Therms	
Commercial	PGE2005	0.45	0.59	0.13	0.47	0.47	0.47	High-Tech Facilities
Facilities	PGE2007	0.80	0.85	0.21	0.60	0.60	0.60	Lrg Com Office Buildings
	Refrig Door Gask - PG&E/SDG&E	0.03			0.19			
	Refrig Door Gask - SCE	0.13			0.19			
	Refrig Strip Curt - PG&E/SDG&E	0.39			0.40			
	Refrig Strip Curt - SCE	0.85			0.40			
Local	<b>Univ of CA / CSU</b>							
Government	PG&E	1.10	1.16	0.59	0.69	0.75	0.72	
Partnerships	SCE	0.38	0.57		0.69	0.75		
	SCG			0.62			0.72	
	SDG&E	0.40	0.11	1.08	0.69	0.75	0.72	
	<b>CA Community Colleges</b>							
	PG&E	0.79	0.59	0.74	0.67	0.69	0.67	
	SCE	0.62	0.40		0.67	0.69		
	SDG&E	0.41	0.42	0.74	0.67	0.69	0.67	
	SCG			0.78			0.67	
	<b>Palm Desert</b>							
	Res CAC Early Retire	1.73	1.69		0.74	0.74		
	Res Refric, Chrg, Airflow				0.76	0.76		
	Com Refrig, Chrg, Airflow				0.70	0.70		
	Other Measures				0.85	0.85		
Major	SCE2517	0.80	0.82		0.59	0.57		Std Perf Contr Prog
Commercial	SDGE3010	0.67	0.66	0.98	0.70	0.68	0.85	Energy Savings Bid Program
	SDGE3025	1.54	1.28	0.33	0.54	0.56	0.43	Std Perf Contr Prog
	SCG3513			0.72			0.54	SCG Bus EE Prog
	SCG3503			0.02			1.00	Educ and Train Prog
Retro-	PG&E	0.45	0.31	0.53	0.80	0.76	0.86	
Commissioning	SCE	0.94	2.07		0.86	0.78	0.91	
	SCG			0.93			0.92	
	SDG&E	1.23	2.60	0.21	0.75	0.75	0.68	
Residential	Furnaces			1.00			0.18	Gross RR is verify only
Retrofit	Clothes Washers – PG&E/SDG&E	1.18	0.00	0.46	0.31	0.31	0.31	
Retrofit	Clothes Washers - SCG			0.79			0.29	
	Dishwashers	1.00	1.00	1.00	0.24	0.24	0.24	Gross RR is verify only
	Gas Water Heaters			0.99			0.65	Gross RR is verify only

**Table 7**  
**Key Parameters from CPUC Evaluation Reports for 2006-2008 IOU Programs**

Evaluation	Component	Realization Rates			Net-to-Gross Ratio			Notes
		kWh	kW	Therms	kWh	kW	Therms	
	Low Flow Showerheads			0.70			0.70	Gross RR is verify only
	Low Flow Faucet Aerators			0.68			0.70	Gross RR is verify only
	Insulation - Attic			1.60			0.27	
	Insulation - Wall			0.38			0.27	
	Refrig Recycling - PG&E	0.58			0.51			
	Refrig Recycling - SCE	0.66			0.56			
	Refrig Recycling - SDG&E	0.49			0.58			
	Room AC - PG&E				0.41	0.41		
	Room AC - SCE				0.36	0.36		
	Room AC - SDG&E	0.34	0.42		0.31	0.31		
	Pool Pump - Single Speed	0.68	2.74		0.32	0.32		
	Pool Pump - Multit Speed	0.47	0.23		0.32	0.32		
	Pool Pump - Reset Agreement	0.11	0.53		0.73	0.73		
	Interior CFLs - SDGE3017	0.59	0.36		0.75	0.75		
	Linear Fluorescents - SDGE3017	1.34	0.04		0.72	0.72		
	Exterior CFLs - SCE2502	0.70			0.75	0.75		
	Interior CFL Fixtures - SCE2502	0.75	0.57		0.77	0.77		
	Interior CFLs - SCE2503	1.30	0.90		0.72	0.72		
	Linear Fluorescents - SCE2502	1.65	1.54		0.77	0.77		
	MF Interior CF Fixtures - PGE2000	0.68	0.36		0.80	0.80		
	MF Ext CF Fixtures - PGE2000	0.84			0.80	0.80		
	MF Interior CFLs - PGE2000	0.60	0.35		0.59	0.59		
	MF Linear Fluorescents - PGE2000	0.12	0.08		0.81	0.81		
	Interior CFL - SDGE3006	0.67			0.44			
	Interior CF Fixtures - SCE2501	0.52			0.66			
Upstream	CFLs	0.24	0.20					Net realization rates
Lighting	Fixtures	0.30	0.86					Net realization rates
	LEDs	0.58	0.00					Net realization rates
PG&E Fab,	Pump-off Controllers	0.46	0.47		0.45	0.44		
Process, & Mfg	All Other Measures	0.53	0.51	0.68	0.60	0.59	0.31	
Small	Interior Screw Lighting - PG&E	0.23	0.24		0.59	0.62		
Commercial	Interior Screw Lighting - SCE	0.21	0.16		0.61	0.64		
	Interior Screw Lighting - SDG&E	0.12	0.07		0.85	0.83		
	High Bay Fluorescent - PG&E	0.63	0.45		0.68	0.68		
	High Bay Fluorescent - SCE	0.69	0.51		0.68	0.70		
	High Bay Fluorescent - SDG&E	0.55	0.58		0.95	0.95		
	Linear Fluorescent - PG&E	0.70	0.73		0.73	0.74		
	Linear Fluorescent - SCE	0.67	0.75		0.79	0.79		
	Linear Fluorescent - SDG&E	0.40	0.48		0.87	0.87		
	Occupancy Sensor - PG&E				0.68	0.70	0.36	
	Occupancy Sensor - SDG&E				0.75	0.60		
	Other Measures - PG&E				0.34	0.43	0.95	
	Other Measures - SCE				0.90	0.87	0.32	
	Other Measures - SDG&E				0.58	0.53	0.01	
Southern	Pipe Insulation - SCG			0.08			0.72	
California	Pipe Insulation - PG&E			0.35			0.49	
Industrial and	Small Com Steam Traps - PG&E			0.30			0.62	
Agricultural	Small Com Steam Traps - SCG			0.12			0.70	
	Small Com Steam Traps - SDG&E			0.12			0.72	

**Table 7**  
**Key Parameters from CPUC Evaluation Reports for 2006-2008 IOU Programs**

Evaluation	Component	Realization Rates			Net-to-Gross Ratio			Notes
		kWh	kW	Therms	kWh	kW	Therms	
	Ind Steam Trap - High Pressure			2.15			0.52	
	Ind Steam Trap - Low Pressure			2.19			0.57	
	Pump Testing - SCE	1.02	0.76		0.84	0.84		
	Other Ind-Ag Measures	0.72	0.65		0.46	0.42		
Specialized	Refrig Chrg and Air Flow - Res	0.15-0.40	0.18-0.40			0.63-0.97	0.63-0.98	
Commercial	Refrig Chrg and Air Flow - C&I	0.09-0.48	0.06-0.75			0.54-0.94	0.54-0.95	
	AC Replacement - Res	0.25-0.46	0.26-0.56			0.55	0.55	
	AC Replacement - C&I	0.47-0.93	0.82-1.12			0.95	0.95	
	Duct Sealing - Res	0.41-0.51	0.41-0.52			0.54-0.96	0.54-0.97	
	Program: SCE2537	0.81	1.10			1.00	0.84	MAP: CO Sensor
	Program: SCE2537	1.00	1.00			0.80	0.80	MAP: Turbocor
	Program: SCE2561	0.58	1.00			0.94	0.95	Ent Ctrs: Dem Cntrl Vent
	Program: SDGE3029	0.95	0.95			0.80	0.80	Upstr HVAC: PTAC/PTHC
	Program: SDGE3029	0.78	0.79			0.98	0.97	Upstr HVAC: HE Motors
Residential	PG&E	1.20	1.19	0.50	1.02	1.03	0.49	
New	SCE	4.83	4.77		0.94	0.97	0.85	
Construction	SCG	1.95	1.95	0.05	0.44	0.45	-0.29	
Nonresidential	PG&E	0.83	0.57	1.20	0.63	0.59	0.83	
New	SCE	1.07	1.12		0.63	0.65		
Construction	SCG			0.70			0.70	
	SDG&E	0.83	0.57	0.67	0.64	0.58	1.22	
PG&E Ag	PG&E	0.70	0.53	1.07	0.70	0.78	0.69	
and Food	SCE			0.63			0.63	
Processing	SCG			0.39			0.46	

Source: CPUC 2006-2008 evaluation reports can be accessed at the following Internet address: <http://www.energydataweb.com/cpuc/default.aspx>

What the above table indicates is that gross realization rates differ substantially from 1.00 in many cases, indicating that the CPUC EM&V activity revealed considerable differences from the IOU ex ante estimates. In addition, the CPUC-estimated net-to-gross ratios (NTGRs) were often fairly low, indicating that many of the IOU program participants would likely have installed measures, anyway, in the absence of the IOU programs.

In addition to summary findings, the CPUC evaluations also report on many key parameters that cause evaluated savings to differ from ex ante assumptions, including factors such as installation rates and unit energy savings. Review of the evaluation reports may be useful in helping the POUs refine program savings estimates.

The recent experience with the CPUC-sponsored evaluations of the IOU programs leads to several conclusions:

- More rigorous program evaluation is useful in assessing the effectiveness of energy efficiency programs, including developing a better understanding of what energy is really being saved and how much of these savings may have occurred anyway. Simple verification of installations and pass-through of unit energy savings and

NTGRs often does not provide enough information to significantly improve one's confidence in program savings accomplishments.

- Gross savings from a number of measures was found to be substantially different from ex ante assumptions, and these results may be useful in adjusting some of the current POU measure savings estimates.
- NTGRs will continue to evolve in a changing program environment and the NTGRs used by the POUs should be revisited regularly. If POU evaluations do not include net-to-gross analyses, the POUs should be reviewing their current NTGR assumptions in light of the recent CPUC evaluation findings.

The CPUC-sponsored evaluations of the 2006-2008 IOU programs used a fairly high degree of rigor for the key programs and measures that represented the majority of expected IOU program savings. Based on the California Evaluation Protocols, the evaluations of the IOU programs mainly used Basic and Enhanced rigor to evaluate gross savings:

- Basic Rigor, used for the less significant programs and measures, involves use of simple engineering models or comparisons of normalized annual consumption that is based on pre- and post-program energy use as taken from utility bills. Basic rigor also involves physical inspection of installation and spot or short-term measurements depending on measure type. Statistically valid samples are required.
- Enhanced Rigor, used for high-impact measures and programs, involves use of: (1) fully specified regression analysis of pre- and post-retrofit consumption, (2) calibrated building energy simulation models as described in IPMVP Option D; (3) retrofit isolation engineering models as described in IPMVP Option B; or (4) experimental design within the program implementation process that compares energy consumption between treatment and non-treatment groups. More extensive monitoring is used to support engineering models, and statistically valid samples are required.

The current POU evaluations use a variety of techniques, ranging from simple verification of measure installation up to enhanced rigor evaluations. However, it appears that the general rigor level of the POU evaluations is lower than that used in the IOU evaluations, which is likely the reason for the substantially higher realization rates, along with the lack of net-to-gross analysis.

The CPUC evaluation protocols also address three levels of rigor in determining net savings:

- Basic Rigor involves analysis of participant self-report survey data.
- Standard Rigor can involve: (1) analysis of participant and non-participant consumption data that addresses the issue of self-selection; (2) an enhanced self-report method using other data sources relevant to the decision to install a measure in addition to participant self report data; or (3) econometric or discrete choice modeling using participant and non-participant comparisons and addressing the issue of self-selection.

- Enhanced Rigor which involves “triangulation” using more than one of the methods in the Standard Rigor level.

For most POU evaluations, net-to-gross issues have not been addressed. To get a better understanding of net program savings, the POUs should use at least the Basic Rigor level (participant self-reports) in their evaluations.

## Reporting Requirements

Evaluation reports should include a minimum level of information that allows the reader to understand the context of the evaluation, understand the methods used in the evaluation, and understand how evaluation findings compare to ex ante program savings estimates.

The CPUC California Evaluation Protocols provide a section on evaluation reporting guidelines<sup>19</sup> that may be overly extensive for POU evaluation report but still give a good sense for the types of information that should be considered for evaluation reports. For energy impact evaluations, the Protocols indicate that the following elements should be included:

- Program ex ante net and gross kW, kWh, and therm savings goals.
- The administrator-generated annual gross kW, kWh, and therm savings (consistent with reported savings).
- Evaluation projected annual gross and net MW, MWh, and therm impacts, measure counts per participant.
- Measure counts versus program goals.
- Measure-level savings.
- Measure reliability metrics (precision levels, coefficients of variation, P-values).
- Savings comparison.
- Appendices discussing the differences between ex ante and ex post results and presenting the weather data used in the evaluation, if any.

The Protocols also provide a typical evaluation outline:

- a. Cover
- b. Title Page

---

<sup>19</sup> See *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*, CPUC, April 2006, pp. 177-203.



- c. Abstract
- d. Table of Contents
- e. Executive Summary
- f. Introduction and Purpose of the Study
- g. Description of Programs Covered in Study
- h. Study Methodology
- i. Reliability Assessment of the Study Findings
- j. Detailed Study Findings
- k. Recommendations for Program Changes
- l. Appendix A – Presentation and assessment of performance metrics
- m. Appendix B – Discussion of the success and timing of the data requests provided to the Administrators

The Protocols state that the Study Methodology section should include the following:

- a. Overview of the approach
- b. Questions addressed in the evaluation
- c. The Protocols and rigor levels assigned to the study
- d. Description of the study methodology
- e. How the study meets or exceeds Protocol requirements
- f. How the study addresses issues presented in the Protocols regarding the methods
- g. Sampling methodology
- h. Expected precision or power analysis results (as required by the Sampling & Uncertainty Protocol)
- i. Sample descriptions (including population characteristics, contact information availability and sample disposition rates)
- j. Description of the baseline
- k. Sources of baseline data
- l. Description of measures
- m. Assumptions on measure performance (including data sources).

Given differences in size and focus between the CPUC-sponsored evaluations of IOU programs and the evaluations of POU programs, we feel that strict adherence to the CPUC

reporting protocols may not be warranted for POU evaluation reports. Some key elements of the Protocols should be considered for each POU evaluation. The elements include:

Program description – This should provide the reader with an understanding of the program and its components that delivered the savings. The description should include program metrics: POU claimed ex ante net and gross kWh, kW, and therm savings (possibly by measure or measure group) and participant counts. The description should also explain how the program fits in with the POUs total energy efficiency portfolio and how savings estimates cited in the evaluation compare to savings provided in annual reports.

Evaluation methodology – This topic should address the:

- Sampling approach, including the method used, such as, simple random sample, stratified ratio sample, target sample sizes, and expected precision levels. If different samples are utilized for gross and net savings estimation, then each sample design should be addressed. The sample design section should clearly show how the sample relates to the program population.
- Data collection methods used (phone surveys, on-site surveys, vendor interviews, literature review) with sufficient detail to provide the reader with an understanding of how the evaluation data were obtained. Survey instruments and sample disposition reporting should be included in appendices.
- Analysis approach used to estimate savings with enough detail for the reader to understand how the analysis was conducted. If an engineering approach is used, a description of the engineering analysis should be provided (using equations, if possible). If a billing analysis is used, a description of the model(s) should be presented that shows structural form and variables included in the analysis.
- Sample expansion approach, which shows the reader how the results of the evaluation sample have been expanded to program-level (or measure-level) findings.

Evaluation results – This piece should present results for the evaluation sample *as well as program-level findings*. This is highlighted as a key area of deficiency associated with the POU evaluations completed to date. Typical evaluation results include elements such as unit energy savings, gross realization rates, net-to-gross ratios, ex post gross and net savings estimates (kWh, kW, therms), and sometimes estimates of key intermediate parameters (such as hours of operation). Precision levels should be reported for key parameters. In addition to the presentation of evaluation findings, a comparison to ex ante estimates should be included. Comparisons should address both gross and net savings estimates, and realization rates should be provided. Comparison can also include key intermediate parameters (such as hours of use, wash load per year, baseline lighting wattages).

Recommendations for program improvements that could involve such items as changes in measure savings assumptions and/or program delivery approaches.



## CHAPTER 5: Conclusion

The EM&V reports submitted by POUs to date represent a first effort to independently verify the savings associated with program measures. As of April 2010, 11 POUs have submitted 13 EM&V studies per AB 2021. Twelve studies were found to include impact evaluations that provided third-party review of claimed program savings, with the remaining study focused on measure persistence and market potential of lighting measures.

The overall goal of the evaluation efforts should be to provide the Energy Commission with an independent assessment across the POUs portfolio of programs of total claimed savings. With the notable exception of SMUD, most of the POUs have had little to no experience with program evaluation prior to these efforts. SMUD, however, has been running efficiency programs for years and has participated in statewide programs with the California IOUs, and that experience showed in the more comprehensive evaluation studies submitted by SMUD.

The following sections summarize the key findings and results of the preceding analysis and application of the Framework to Evaluate EM&V studies. Recommendations are provided to improve the rigor, documentation and consistency of EM&V efforts to support the SB 1037 and AB 2021 requirements.

### Key Findings and Recommendations

#### Timing of EM&V and SB 1037 Reporting Requirements

One area of confusion appears to be the relationship between the EM&V reports and the SB 1037 annual reports. For the most part, it appears that POUs are using the SB 1037 annual report to present the claimed (ex ante) portfolio savings based on their completed EE Reporting Tools, and are not presenting the evaluated (ex post) savings. The one exception was Truckee Donner, which reported its evaluated savings in the SB 1037 report.

The SB 1037 annual reports have a submittal deadline of March 15 for the previous program year, which for most POUs runs on a fiscal year basis from July 1 through June 30. This generally provides six to seven months for evaluation work, which is a relatively tight timeline. A few POUs are operating energy efficiency programs based on the calendar year, which means the programs would not be completed until December 31. In these cases, it is not feasible to report the evaluated savings with the March 15 SB 1037 submittals.

#### *Recommendations*

Due to the range of program year completion dates, it is not feasible for all POUs to report evaluated (ex post) savings within the March 15 SB 1037 annual reports. Therefore, it is recommended that a regular reporting deadline for EM&V reports and verified savings be established. This may be implemented in a few ways:

- **Separate EM&V deadline from SB 1037 deadline**, with a different deadline for POUs with programs based on fiscal year and POUs with programs based on calendar year. A minimum of seven to eight months after program completion is

recommended for quality evaluation results, assuming an evaluation contractor is selected and in contract before the end of the program year.

- **EM&V included with SB 1037 deadline.** Potentially, the evaluation results could be included in the following years SB 1037 report. The FY2007/08 evaluation results summary included in March 15, 2010 along with FY2008/09 claimed ex ante savings).

### *SB 1037 Documentation Adequacy*

The SB 1037 annual reports submitted for 2006, 2008, and 2009 include only net savings values. This contributes to confusion relative to the EM&V reports, which, in many cases, also lack clear information related to whether savings are gross or net values. The SB 1037 report should specify both gross and net savings to promote transparency in the net-to-gross assumptions. The portfolio level savings should also be broken down by program, in addition to the measure categories, to aid in the understanding of the EM&V reports.

### *Recommendations*

Require SB 1037 annual reports to include the following information:

- Clear indication of whether reported savings are claimed (ex ante), or whether evaluated (ex post) savings are being reported.
- Savings tables from the “Results” tab of the EE Reporting Tool (See **Figure 22** for example of the table) to report savings both by program and by measure category. The table also includes both gross and net claimed (ex ante) savings values

The Energy Commission should also request the completed EE Reporting Tool from the POUs as documentation of their SB 1037 claimed (ex ante) savings, as well as calculations supporting the revision of default values in the EE Reporting Tool or addition of any custom measures.

### *Contextual Reporting Adequacy*

In general, the POU evaluation studies committed the greater part of the evaluation to the one or two programs that were responsible for the majority of portfolio savings, but the exact percentage was usually not reported. A supplemental process evaluation, or scaled down impact evaluation, was then used to evaluate a program or two of smaller savings. The remaining portfolio savings were usually not mentioned at all, nor was there much discussion of the relative uncertainty of the reported savings. Therefore, the POU evaluations appear to fall short of providing complete descriptions of the fundamental reporting elements typically expected in an EM&V report.

While the EM&V reports often included a list and description of the POUs portfolio of efficiency programs, information was lacking in the report to fully understand the savings being evaluated and how the results compare with the POUs claimed ex ante savings at the portfolio level. This context is needed, especially to understand what the verified portfolio level savings would be.

## *Recommendations*

Provide guidance to POUs requesting that the following components be included in future EM&V reports:

- Providing a more complete program description and associated ex ante savings estimates (matching SB 1037 reported values).
- Explaining how the evaluation fits in with the program and how the program fits in with the POU program portfolio (as documented in the SB 1037 annual report).
- Describing rationale for the programs (or components of the portfolio) selected for evaluation, including a discussion of the relative uncertainty of savings associated with different programs and components of the portfolio.
- Documentation of evaluation methods, especially sample design.
- Explaining the expansion of evaluation results to the program level.
- Comparing ex post evaluation results with ex ante savings estimates.

For POUs that contract out evaluation services separately for different efficiency programs, a summary report should be included that provides the portfolio level verified savings, indicating which measure savings or programs were not evaluated.

## **Gross Savings Approach**

The POU evaluations differ from the IOU evaluations most significantly in scope and rigor. The evaluations were heavy on installation verification with a review and critique of the deemed savings methods. Primarily this was used for residential appliance rebates and non-residential lighting retrofits. For appliance rebates, the verification and deemed savings review is appropriate. For other non-residential projects, however, this is generally not sufficient. The incremental cost of adding time-of-use logging and/or spot measurement to on-site verification is small enough that it is easily justified by the added confidence in the estimate. This is especially true when these projects account for a large percentage of the POU savings portfolios and that most of these programs have not been evaluated previously. Several of the evaluation studies did not meet the basic level of rigor required by the California Evaluation Protocols.

In most cases, the methods and algorithms used to calculate evaluated gross savings and verification results were not clearly documented. In a robust evaluation report, the documentation should include all of the necessary details, such as assumptions and methodology such that the reader does not have to guess at the assumptions made and reverse-engineer calculations. A properly documented evaluation should be comprehensive enough that another reasonably competent consultant could recreate the analysis with another dataset. This does not mean the evaluator needs to share every line of code produced for the analysis; it means that the approach is well-defined and described clearly in the report, and that all of the methodological steps and calculations are clearly stated.

Unfortunately, insufficient and partial documentation was the case in most of the POU evaluations. While basic approach and some explanation were stated, the pathway from

raw data to the final results was not clear in most cases, including baseline assumptions for the energy savings estimate. A considerable number of evaluations gave few clues beyond general site observations, and no clear explanation of parameters adjusted and how the ex post savings were calculated.

## Recommendations

In the effort to improve the quality and transparency of evaluated gross savings estimates, the POU evaluations should:

- Evaluate the programs representing a significant share of portfolio savings, at minimum at the basic level of rigor, either through site specific revised savings estimates or billing analysis.
- Address EUL at a minimum, adjust any measure life assumptions, and as best practice should provide a table that shows the lifecycle savings estimate over a given period of years (out to the longest measure life in the program).
- Provide tables that show ex ante measure assumptions compared with field results. Parameters may include fixture counts (quantity), wattage, operating hours, set-point, temperature, and kW consumption.
- Include on-site data collection forms and telephone survey instruments in the EM&V report.
- Review the baseline assumptions associated with the deemed savings and interview program participants during the verification visits to verify baseline assumptions, such as fixture wattage previously installed and remaining useful life of equipment that was replaced prior to its end of life.

## Sampling Approach

Reviewers found a number of issues that seemed to stem from a lack of statistical knowledge and experience among both the evaluators and the POUs, including missing discussion of reliability of results, statistical formulas misapplied, and precision estimates that did not pass the “sniff test.” While most studies showed an awareness of the need for statistical rigor, they fell short in applying statistical methods to develop a sample plan, and especially in evaluating and presenting the reliability of the results of the evaluation. (The sampling precision targets are necessary to assess how reliable the resulting savings estimates are.)

The goal of sampling and research design is to estimate actual gross savings *for the overall program* in a sound, defensible, and unbiased way. If a proper sample design was used, then the results of the sampled projects can be used to develop estimates of program population results that will be close to the true values that would have resulted if the same data collection had been carried out for all projects. Many of the EM&V studies did not make any attempt to complete the final step to expand the evaluation results of the sample back to the program population. These studies provided realization rates only for the sampled sites but did not provide any meaningful results at the program level.

### *Recommendations*

- A specified confidence interval and desired relative precision should be used to determine the size of the sample necessary to meet those targets.
- Where precision is calculated from multiple evaluation study efforts, the precision information should be provided for each study effort as well as the combined result.
- The report should document both the desired relative precision and the achieved precision.
- The evaluations should provide program level effects resulting from the EM&V work. At a minimum, the program savings should be adjusted to reflect the verified savings associated with the sampled projects, with an explicit note that the remainder of program savings were passed through from the ex ante claim.

### **Net-to-Gross Analysis**

Another key weakness in the POU evaluation efforts is the lack of net-to-gross assessment. Although many of the EM&V reports mentioned net-to-gross, only a few studies developed net-to-gross ratios to adjust for net savings. The California IOUs and CPUC have developed relatively robust net-to-gross methodologies, including standardized self-report survey questions and analysis. The POUs could easily adopt this approach for their EM&V efforts.

### *Recommendations*

- Leverage the IOU net-to-gross methods to ensure consistency in results.
- At minimum, the EM&V reports should either complete an adjusted net-to-gross ratio analysis for the program, or clearly indicate a “pass-through” of the deemed net-to-gross ratio is being used to estimate net savings.

### **Guidance for POU Evaluation Efforts**

Overall, it appears that POU evaluation funding is on the low side and may need to be increased if POU EM&V activities are to keep pace with IOU evaluation activities. CPUC set EM&V budget at 8 percent for the 2006-2008 IOU energy efficiency programs, and then at 4 percent for the 2010-2012 program years. POU budgets maybe should be on the higher side relative to IOU evaluations because of the small size of POU programs and the diversity of program delivery processes. Some POUs might consider combining EM&V activities to gain some economies of scale, with this being especially appropriate for POUs with similar efficiency programs (both measure type and program delivery processes).

POU staff needs more guidance and training on the expected rigor, activities, and scope of the evaluation efforts to ensure the reliability of the verified ex post savings. Some statistical training or primer for evaluators would also be useful. POU staff writing RFPs and managing evaluation projects need to have sufficient statistical knowledge to interpret and review the work being performed by the evaluator.



## Recommendations

The Energy Commission may consider the following approaches to help educate POU staff:

- **Budget Guidelines.** Consider recommending that POUs set EM&V budgets in the 4 to 8 percent range, consistent with recent CPUC guidelines.
- **POU EM&V Handbook.** Develop an EM&V guidance document for POUs to detail and explain expectations. The document should leverage the 2004 *California Evaluation Framework* and 2006 *Evaluators' Protocols* but provide specific examples and pare down the content to be applicable to POU programs. The handbook may the following sections:
  - Gross savings methods, including both engineering and billing analysis
  - Net-to-gross methods
  - Sampling and statistical precision
  - EM&V reporting requirements
- **Trainings and Workshops.** In-person trainings or webinars would be useful for explaining the concepts and requirements included in the POU EM&V Handbook. An in-person introductory training is recommended, with webinars and delving into specific topics and sections of the handbook.

## EM&V Reporting Requirements and Documentation

To improve transparency and consistency of EM&V reports, POUs should provide more documentation associated with evaluation efforts. For instance, although some evaluation proposals and pricing bids were submitted to the Energy Commission through the Northern California Power Agency (NCPA) website, no information was provided on actual budgets spent. This information would be useful to assess how effectively evaluation dollars are being spent and whether additional funds would improve the quality of evaluations.

The POUs may also benefit from a consistent reporting template for EM&V studies. Furthermore, more complex studies require clear report organization and explanation for the reader to follow the often complicated methods and analysis in the study, as some of the longer POU EM&V reports fell short in that area.

## Recommendations

The following reporting templates and guides may also assist POUs to include the contextual and methodological documentation to aid in the review of the EM&V reports.

- Example EM&V Report Outline
- Standard EM&V Results Reporting Table (both annual and lifecycle savings table)
- Example Checklist for POU EM&V report

**Figure 21: Example Checklist for POU EM&V report**

<p><b>Contextual Reporting</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Does the EM&amp;V report clearly state savings values consistent with the associated SB 1037 annual report?</li> <li><input type="checkbox"/> Does the evaluation cover a significant portion of the POUs portfolio and clearly describe the programs or savings not evaluated?</li> <li><input type="checkbox"/> Does the evaluation assess risk or uncertainty in selecting the components of the portfolio to evaluate?</li> </ul> <p><b>Overview and Documentation of Specific Evaluation Effort</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Does the report clearly identify what is being evaluated in the study (part of a program; an entire program; the entire portfolio)?</li> <li><input type="checkbox"/> Does the evaluation include an assessment of EUL and lifecycle savings?</li> <li><input type="checkbox"/> Does the evaluation report provide documentation of all engineering and billing analysis algorithms, assumptions, survey instruments and explanation of methods?</li> <li><input type="checkbox"/> Does the report describe the methods in sufficient detail that another evaluator could replicate the study and achieve similar results?</li> <li><input type="checkbox"/> Are all data collection instruments included, typically in an appendix?</li> <li><input type="checkbox"/> Does the report adequately describe metering equipment and protocols, if any, typically in an appendix?</li> </ul> <p><b>Gross Savings</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Does the report review the program's choice of baseline?</li> <li><input type="checkbox"/> Does the report clearly characterize the population of participants?</li> <li><input type="checkbox"/> Does the report clearly discuss its sampling approach and sample design?</li> <li><input type="checkbox"/> Does the report state the sampling precision targets and achieved precision?</li> <li><input type="checkbox"/> Does the report clearly present ex post savings?</li> <li><input type="checkbox"/> Are the results expanded to the program population? If not, the report should state why not and clearly indicate where ex ante savings are being passed through.</li> <li><input type="checkbox"/> Does the study clearly explain any differences between ex ante and ex post savings?</li> </ul> <p><b>Net Savings</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Does the evaluation include a quantitative assessment of net-to-gross? If not, does the evaluator clearly indicate the source of the assumed net-to-gross value?</li> <li><input type="checkbox"/> Does the report clearly discuss its sampling approach and sample design?</li> <li><input type="checkbox"/> If a self-report method is used, does the approach account for free-ridership?</li> </ul> <p><b>EM&amp;V Summary and Conclusions</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Does the report provide clear recommendations for improving program processes to achieve measurable and cost-effective energy savings?</li> <li><input type="checkbox"/> Does the evaluation assess the reliability of the verified savings and areas of uncertainty?</li> </ul>
--

Source: *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals* ("Evaluators' Protocol"). Prepared for the California Public

Utilities Commission. April

2006. [http://www.calmac.org/events/EvaluatorsProtocols\\_Final\\_AdoptedviaRuling\\_06-19-2006.pdf](http://www.calmac.org/events/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf)

*The California Evaluation Framework*. Prepared for the California Public Utilities Commission and the Project Advisory Group. June

2004. [http://www.calmac.org/publications/California\\_Evaluation\\_Framework\\_June\\_2004.pdf](http://www.calmac.org/publications/California_Evaluation_Framework_June_2004.pdf)

## Conclusion

A study can be perfectly planned and executed but can still be ineffectual if the results are reported in a manner that is confusing, unclear, or incomplete. Most of the POU EM&V studies had issues with how the results of the impact evaluation were reported. A number of studies seemed to lose sight of the goal of an impact evaluation, that is, to provide an estimate of the actual savings of a program.

There are significant California specific resources available to support POU EM&V efforts, from the CPUC-sponsored 2004 *Evaluation Framework* and 2006 *Evaluators' Protocols*, to standard net-to-gross survey questions, to CPUC Evaluation Reporting Protocols. As POUs gain experience with evaluations, it is expected that evaluation efforts will continue to improve and develop to promote more consistency across state-wide estimates of energy efficiency program impacts.

## Glossary

Acronym	Definition
AB 2021	Assembly Bill 2021
CMUA	California Municipal Utilities Association
CPUC	California Public Utilities Commission
DEER	California database designed to provide well-documented estimates of energy and peak demand savings values, measure costs, and effective useful life for selected EE technologies and resources. See <a href="http://www.deeresources.com">www.deeresources.com</a>
DEEMED SAVINGS	Estimate of energy savings for an installed energy efficiency measure from data sources, such as DEER, or other documents acceptable analytical method.
EM&V	Evaluation, Measurement, and Verification
Ex-Ante Savings	Ex-ante savings are the reported savings calculated with deemed savings figures prior to EM&V
Ex-Post Savings	Ex-post are the adjusted savings based upon the results EM&V studies
Free Ridership	Program participants who would have implemented the program measure or practice in the absence of the program. Free riders can be total, partial, or deferred
Gross Savings	The change in energy demand that results directly from program-related actions taken by participants in an efficiency program
GWh	Gigawatt hour
HVAC	Heating, ventilation, and air conditioning
IOU	Investor-owned utility
Kw	Kilowatt - Unit of power equal to 1,000 watts
Kwh	Kilowatt Hour - Unit of energy equal to 1000 watt hours
MMth	Million therms
MW	Megawatt - Unit of power equal to one million watts
Net Savings	Change in electricity demand that is attributable to an energy efficiency program.

Net-To-Gross Ratio (NTG):	A factor representing net program savings divided by gross program savings that is applied to gross program impacts to convert them into net program load impacts.
NCPA	Northern California Power Agency
POU	Publicly owned utility
SB 1037	Senate Bill 1037
SCPPA	Southern California Public Power Authority
TMG	Total market gross
TRC	Total resource cost



**APPENDIX A:**  
**Summary of Each EM&V Study Assessed**





**Table A1. Summary of the POU EM&V Reports**

<b>POU EM&amp;V Report</b>	<b>Evaluated Program Year</b>	<b>Evaluation Report Completion Date</b>	<b>Programs Evaluated for Impacts</b>	<b>General Approach</b>
Alameda Municipal Power	FY 2007/08	6/8/2009	Commercial Custom Program	5 largest rebated projects – on-site verification and site specific savings re-calculation
Lodi Electric Utility	FY 2007/08	12/2008*(Taken from Document Properties)	Non-residential Custom Program	All 5 lighting projects - on-site verification and site specific savings re-calculation
City of Lompoc	FY 2008	3/10/2009	Refrigerator Rebate Program Refrigerator BuyBack Program Income Qualifying Refrigerator Purchase Program	Sample of 21 replaced residential refrigerators - Paper verification and compared deemed savings to ENERGY STAR website deemed savings
City of Palo Alto Utilities	FY 2007/08	2/19/2009	Residential Refrigerator/Freezer Recycling	Paper verification (review of database) and review of deemed savings
			Residential CFL Program	Sample of 50 participants for telephone verification; and review of deemed savings
			Right Lights Program	20 sampled projects – on-site verification and site specific savings re-calculation
			Non-residential Custom Program	6 sampled projects – on-site verification and site specific savings re-calculation
Port of Oakland	FY 2007/08	2/2/2009	Non-residential Custom Program	1 site (census) – on-site verification and site specific savings re-calculation
Redding Electric Utility	FY 2007/08	3/25/2009	EarthAdvantage Program	70 sampled sites for paper verification. 1,252 sampled sites (census?) – billing analysis for impacts by measure category
Roseville Electric	FY 2007/08	2/27/2009	Residential New Construction	57 applications (out of 315) - Paper verification and discussion of relative merits of deemed savings estimates
			Residential HVAC Retrofit	57 applications (out of 350) - Paper verification and discussion of relative merits of deemed savings estimates

POU EM&V Report	Evaluated Program Year	Evaluation Report Completion Date	Programs Evaluated for Impacts	General Approach
			Commercial Custom Program	21 applications (census) - Paper verification and discussion of relative merits of deemed savings estimates
Silicon Valley Power	FY 2007/08	3/20/2009	Non-residential programs	10 sampled projects (out of 147) – on-site verification, spot measurements, 1-2 week metering period
SMUD	CY 2006	5/1/2007	Refrigerator Recycling Program	Participant telephone survey and in-situ monitoring
	CY 2006/2007 <sup>1</sup>	3/31/2008	Residential HVAC Program	60 sampled participating homes – No verification, but metering to compare participant to non-participant energy use
Truckee Donner (TDPUD)	CY 2008 <sup>2</sup>	2/20/2009	All 17 programs evaluated	Telephone survey and on-site verification, monitoring
Turlock Irrigation District	FY 2008	3/16/2009	Non-residential rebate program	2 sampled sites - on-site verification, 1 week monitoring, and site specific savings re-calculation

---

<sup>1</sup> The savings estimates for the study were reported on a per-unit basis and not extrapolated to the total participant population. Data was collected in 2007, and included installations from 2006 and the beginning of 2007.

<sup>2</sup> It was not clear from report whether this is calendar year 2008 or fiscal year 2008, but the March 15, 2009 SB 1037 annual report indicates that the TDPUD programs are calendar year.

## City of Alameda (Alameda Municipal Power)

**Brief Program Description** – The report focused on the Commercial Retrofit Program. The report indicates that program measures include lighting, controls, and motors.

**Brief Evaluation Description** – A telephone survey of program participants and non-participants was completed, primarily for a process evaluation to identify major barriers to program participation and identify areas for program improvement. The impact evaluation included a site visit of four of the five largest projects, with the largest project (Coast Guard facility) excluded due to lack of access.

**Table A2 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		NP	
Gross kW	NP		NP	
Gross therms	NP		NP	
Net kWh	2,135,449.00		NP	
		Net demand savings = 360, Net peak demand = 180		
Net kW	360.00		NP	
Net therms	NA		NA	
Lifetime kWh	NP		NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NA	
TRC	6.21		NP	

**Table A3 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	NP	
Gross kWh	405274	332,575.00	Report provided ex post gross for the 5 site sample
Gross kW	34.8	34.10	Report provided ex post gross kW for the 5 site sample, but "demand" is not defined
Gross therms	NA	NA	
Net kWh	NP	NP	
Net kW	NP	NP	
Net therms	NA	NA	
TRC	NP	NP	
Realization rate kWh		0.82	
Realization rate kW		0.98	
Realization rate therms		NA	

**Table A4 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	1	EM&V Reported Portfolio savings seem too high
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	1	Alameda has a very low percentage of 19%, because one large project that accounted for a majority of the portfolio savings did not allow any verification or evaluation activities. Otherwise, Alameda did evaluate the program that included almost all of the savings for the portfolio. Therefore, without this one problem site in the population, the evaluation would have covered a majority of the portfolio savings. Additionally, the large site did not receive any monetary incentive, so its inclusion in the program and associated reported savings is open for debate.
Percent of portfolio covered (estimated)	%	100%	
Gross kWh of all programs evaluated	kWh	2,420,400	Best guess, not clear
Net kWh of all programs evaluated	kWh		
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	1	
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	3	Because Alameda had a very small portfolio, this report was easily able to evaluate a large percentage of the portfolio
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	2	Alameda completed a billing analysis and verification with deemed savings review and self-reported hours. The billing analysis was completed at the basic rigor level, using whole facility analysis of utility meter data for the baseline and reporting period.
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	EUL was not mentioned or addressed by the report
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	2	Partial explanations were provided

Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	3	Claimed on-site, but did not report approach
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	Alameda completed an on-site verification with site-specific adjustments to annual operating hours based on facility personnel interviews.
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	2	Alameda consistently used early replacement baseline without explanation
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	2	Alameda was given a "partially appropriate" score, having surveyed four of eight participants. The evaluation had to omit the largest project in the program due to lack of customer cooperation, but the project was so large that without it the sample could not be considered appropriate.
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	1	
Precision target		NP	No target was reported
Precision achieved		NP	It was not mentioned whether precision was achieved or not
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	2	The Alameda billing analysis utilized a very coarse average daily temperature approach by month rather than a more precise and accepted cooling degree day approach. Furthermore, the results were extrapolated from just a few months of post-implementation data that produced an unrealistic trend line. For Site 2, the evaluation contractor found that while the calculated savings were more accurate (based on the site's actual operating hours), the deemed savings resulted in a higher savings number and subsequently the report recommended that the site claim the deemed savings value. This is not an acceptable verification of project savings.
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	1	Sample results were not extrapolated
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	2	Alameda provided some discussion of discrepancies in narrative (paragraph) format, however the discussion is hard to follow and savings table would made it great deal easier to understand
Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	2	The evaluation used self-report surveys to explore free-ridership and offered recommendations based on the findings, but failed to develop a net-to-gross ratio (NTGR) or net savings
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	1	Alameda's evaluation was deemed not to have employed an appropriate sampling approach, since it asked four out of the total five large commercial participants about free-ridership. However, the evaluation report chose not use the survey to generate site or program net-to-gross ratios, citing "insufficient data."
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	1	The Alameda study did not apply the NTG analysis method appropriately, despite the fact that some free-ridership questions were asked of the four out of five large commercial participants during on-site visits. The EM&V report made no adjustments to the gross savings estimate citing the "qualitative nature" of these findings. The specific questions were not included in the EM&V report as documentation.
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	1	Alameda also included a qualitative assessment of free-ridership. Since no discussion of partial free-ridership was included in the report (nor were the survey questions), we presume that partial free-ridership was not addressed.

Section 5 - M&E Summary	Scoring Metric	Score	Notes
Criterion 5.1. Does the evaluation provide clear recommendations?	(1-3)	2	Some process recommendations were included
Criterion 5.2. Does the report assess the reliability of the ex post savings results?	(1-3)		
Overall score		34	

### Findings and Recommendations-

The report did not include much contextual information about the program. Information that would have been useful about the Commercial Retrofit Program, include eligibility (e.g. kW size, types of commercial facilities), size of incentives, types of measures included in the program, including prescriptive or custom measures).

The telephone survey included some free-ridership questions, but there was no net-to-gross methodology for using the answers to determine net-to-gross. The report states that the data is too qualitative and not sufficient. In the future, a more rigorous net-to-gross methodology would improve the evaluation of the program net savings estimate.

The impact evaluation excluded the Coast Guard station, which accounted for 76% of the utility's claimed savings, but for which no rebates were issued. The evaluator should have raised questions about why the savings were claimed if no rebates were issued. The evaluation should also have reviewed any program application or documentation associated with the project to provide an assessment of the validity of these savings (including whether sufficient documentation is provided to even make that assessment).

The documentation of the evaluation of the sites is found to include a narrative description of the findings. The report would have been strengthened with the inclusion of engineering algorithms and comparison of ex ante and ex post assumptions. The narrative would then describe where the field visit found discrepancies between the ex ante and ex post values (e.g. fixture counts, wattage assumptions, operating hours, etc). This information is best presented in the table.

The report also provides a comparison of kW and kWh claimed against the verified calculated savings. It is unclear whether these are net or gross values, and should be specified.

## City of Lodi (Lodi Electric Utility)

**Brief Program Description** – The Lodi energy efficiency programs are organized across four customer types: residential, schools, low-income residential, and commercial/industrial. The impact evaluation focused on two specific programs: Non-residential custom program (lighting measures) and the Residential Appliance Rebate program. The Non-Residential Custom Rebate program provides rebates for qualifying projects. The Appliance Rebate Program provides rebates to customers who purchase ENERGY STAR refrigerators, dishwashers and front-loading clothes washers.

**Brief Evaluation Description** – The evaluation provided a high level process evaluation of Lodi’s implementation of its portfolio of programs, including a review of the residential database to streamline program reporting, and a review of the measures included in the residential programs. The process evaluation outlined program activity flow chart and findings from staff interviews.

For the impact evaluation of the non-residential custom program, the evaluators focused on all five projects with lighting measures and conducted site visits to verify installations and savings assumptions (e.g. operating hours, wattages, etc). Verified savings were provided by adjusting engineering assumptions based on the results of the site visit.

The impact evaluation of the residential appliance rebate program encompassed a paper verification of a sample of application forms to check for completeness, including invoices and sufficient appliance information (e.g. type of appliance, water heating fuel type, baseline appliance age, etc). Verified savings were estimated by adjusting deemed savings based on the results of the paper verification; that is, where no fuel type was included in the application, the evaluators applied the more conservative deemed savings value (e.g. gas versus electric water heating). A 95 percent verification rate was applied to the final savings values across all appliance categories since 5 percent of the applications could not be found.

**Table A5 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		2,908,248.00	Only lighting, claimed to account for 72%
Gross kW	NP		412.40	Only lighting, claimed to account for 72%
Gross therms	NA		NA	
Net kWh	3,090,527		NP	
Net kW	463.0		NP	
Net therms	NA		NA	
Lifetime kWh	34,716,425		NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NA	
TRC	5.92		NP	

**Table A6 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	NP	
Gross kWh	2,908,248	2,772,851 and 12,574	Provided values for lighting and residential appliances
Gross kW	412	378.9, 3.7	Provided values for lighting and residential appliances
Gross therms	NA	NA	
Net kWh	NP	NP	
Net kW	NP	NP	
Net therms	NP	NA	
TRC	NP	NP	
Realization rate kWh		95%	For lighting
Realization rate kW		92%	For lighting
Realization rate therms		NA	

**Table A7 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	3	Report includes gross savings which seem reasonable relative to the SB 1037 net savings. Impacts are evaluated for the NR lighting retrofit and residential appliance program.
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	2	Partially covers Lodi's portfolio of programs
Percent of portfolio covered (estimated)	%	96%	Non-residential lighting is reported as ~72% of the portfolio and the appliance program is reported as ~ 0.3% of the portfolio
Gross kWh of all programs evaluated	kWh	2,785,426	
Net kWh of all programs evaluated	kWh	NP	Report provides no net savings analysis
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	2	Risk and uncertainty implied but not mentioned
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	2	The report had impact evaluations for non-residential lighting retrofits and residential appliance components that accounted for over ~72% of portfolio savings.
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	1	The evaluation performed verification only. On-site to verify lighting, no spot measurements were made. Paper verification was implemented for residential appliance program.
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	EUL was not mentioned or addressed by the report
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	2	Site-by-site explanation of how data was collected, but no explanation of analysis or engineering equations were used.



Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	3	Lighting on-site verification, paper verification of the appliance program, and no other programs
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	3	
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	Yes, sampling approach was reasonable - sampled all 5 sites in population
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	NA	Not applicable - No precision estimates associated with a census
Precision target		NA	Sampled all 5 sites in population
Precision achieved		NA	Sampled all 5 sites in population
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	3	The verification and deemed savings review appears to have been well-executed. However, a savings table associated with each lighting site clearly indicating revised counts and wattages would have cleared up many ambiguities
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	3	Lodi was included as having extrapolated the five lighting sites results to program-wide, but the five lighting sites were a census for that measure category, and there were other measure types included in the Non-residential Custom Rebate Program (although project count and savings were not provided in the report).
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	3	Yes, on the site level, the report discusses the reasons for discrepancies and the implications they have on savings. Site savings table would have been better as the details are hard to follow.
Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	1	EM&V report did not include NTG. No mention of net savings
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NA	No mention of net savings
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	NA	No mention of net savings
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	NA	No mention of net savings
Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	3	Yes, the report has many suggestions for residential measures, and recommendations for the non-res lighting component
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)	1	No assessment of reliability
<b>Overall score</b>		39	

## Findings and Recommendations–

The portfolio was dominated by the five verified lighting sites. Given this bulk of savings, the decision to focus a limited study on the individual site visits appears appropriate. Fixture counts were verified and inquiries were made to site facility staff on hours of operation to check against implementer reported hours of use. Although these activities improved the reliability of the estimates, these estimates could be further improved substantially with two small tasks: time-of-use (TOU) logging and a decision maker survey for these five projects, along with some additional analysis for the data these tasks will provide.

The single greatest risk to these savings estimates is the specter of free-ridership (i.e. that some or all of these projects would have been installed exactly the same absent any program

influence). Free-ridership is prevalent in large facilities, especially among large corporations that operate a number of facilities. Often, these large corporations employ “incentive hunting” consultants with the sole responsibility of finding efficiency program incentive monies for projects that have already been decided upon. Needless to say, the program in these cases has no “net impact” as the project would have been implemented without any program influence.

The assessment of free-ridership is typically assessed via a decision-maker survey where the primary decision-maker for the project is asked a battery of questions that probe for the influence of program. Not only does the decision-maker survey assess blatant free-ridership as discussed above, but also partial free-ridership.

For the five project examined by the evaluation, the baseline used in all cases was the replaced equipment, the so-called early replacement baseline. There was no discussion of baseline selection to be found in the report. Additional estimate certainty can be attained with a baseline assessment and/or evaluation of the “remaining useful life” (RUL) of the replaced equipment.<sup>3</sup> Usually the assessment of baselines and RUL of the equipment can be conducted in the same decision-maker survey as the net-savings assessment, thereby reducing costs.

Also short term time-of-use (TOU) data logging, spot measurements could also add considerable certainty to the savings estimates. Self-reported hours of operation can be improved upon through the use TOU data loggers that record the on and off times of individual fixtures. Depending on the circuit configuration and control strategy, a relatively small number of TOU loggers can be used to either provide a census of lighting operation or, at minimum provide a representative sample of lighting operational data to expand to the entire facility. In some cases a single true power logger can installed on the lighting panel feeder can accurately the lighting usage for an entire facility. There is some uncertainty expanding the monitoring period to an entire year, but short-term monitoring is still considered more reliable than self-reported hours. These efforts are relatively inexpensive, especially when only the increase in cost from a verification and deemed savings review is considered.

The residential program, although a small contributor to total savings, services a significantly different energy efficiency strategy than non-residential. This evaluation only skimmed the surface of its impact by identifying the presence or absence of check receipts. Despite the lesser

---

<sup>3</sup> When a project influences replacement of equipment that would have remained in place and operating for several years, then the early replacement baseline or preexisting equipment is the correct method of the calculating the savings estimate. However, if the equipment is at the end of its useful life, the only savings being realized are due to efficiency of the installed equipment above current code standards. The normal replacement baseline in these cases is equipment or materials are that are minimally compliant with current municipal, state, or federal codes. In cases where the program has influence the timing of the installation, the project is best estimated with a “dual baseline”. A dual baseline uses early replacement baseline until the point where the normal replacement would have occurred, and the normal replacement baseline from that point forward. Typically, if a decision-maker indicates the normal replacement would have occurred past a certain cut-off point, usually three years, the uncertainty of the future implementation is great enough that the early replacement baseline is used for the life of the measure, not a dual baseline.

impact of the residential program, the assumption that savings incurred due to a completed application and mailed check are equivalent to assumed deemed savings only results in minimal improvement over the program's ex ante savings estimates. The reliability of the evaluated savings could have been improved by conducting on-site visits with a handful of sites to physically verify equipment, identify whether equipment was additional or a replacement on burnout, conduct an interview to gather hours of operation data, and calculate rough estimates of installed CFLs.

As a final note, although the report states that customer names are not given due to privacy concerns, the subsequent narrative and tables identifies each site by customer name. This is not typical standard practice for an evaluation report, especially considering that there is more concern over confidentiality than ever before.

## City of Lompoc

**Brief Program Description** – Three refrigerator and freezer programs were grouped together for this evaluation. The programs constituted over 75% of the utility’s claimed energy savings, and include:

- Replace existing refrigerators/freezers and dispose of the replaced appliance
- Purchase existing second refrigerators/freezers and dispose
- Replace existing low income refrigerators/freezers and dispose of the replaced appliance

For the two programs that include replacing the existing refrigerators/freezer, the new appliance are required to be ENERGY STAR.

**Brief Evaluation Description** – Impact evaluation included a paper verification of the documentation associated with the savings claimed by the utility. The documentation was found to be complete. The evaluation also focused on reviewing the deemed savings estimate associated with the refrigerator measure and the freezer measure that was used for each program. The evaluation estimated energy use associated with the refrigerators and freezers removed, by looking up the model type and serial number information in the ENERGY STAR database. These savings were provided as improved savings estimates over the ex ante values.

**Table A8 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		NP	
Gross kW	NP		NP	
Gross therms	NP		NP	
Net kWh	304,163.00		NP	
Net kW	61.00		NP	
Net therms	NA		NA	
Lifetime kWh	NP		NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NA	
TRC	4.41		NP	

**Table A9 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	NP	
Gross kWh	264096	263,703.00	Calculated
Gross kW	1039.7	NP	
Gross therms	NA	NA	
Net kWh	NP	NP	
Net kW	NP	NP	
Net therms	NA	NA	
TRC	NP	NP	
Realization rate kWh		1.00	
Realization rate kW		NP	
Realization rate therms		NA	

**Table A10 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	1	EM&V Reported Portfolio savings seem too high
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	2	
Percent of portfolio covered (estimated)	%	75%	Best guess, not clear
Gross kWh of all programs evaluated	kWh	264,096	Calculated from SB1037 using 75%
Net kWh of all programs evaluated	kWh	NA	Results of M&E report not extrapolated to portfolio level
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	2	The report reduced much of the risk by surveying the largest programs

Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	2	Three similar programs made up 75% of the portfolio
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	1	Verification plus deemed savings assessment
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	EUL was not mentioned or addressed by report
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	1	The Lompoc evaluation did not include any discussion of the algorithm for estimating savings. Although the report explained the steps, there was no description of the overarching methodology

Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	1	Comparison of invoices to database, looked at name plate data and neglected results
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	The Lompoc study collected energy use data for a sample of specific refrigerators recycled by the program from secondary sources, and compared them to deemed savings.
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	2	Did not consider normal replacement or RUL
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	Sampling meets basic rigor
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	2	
Precision target		80/20, 90/15	80/20 by program, 90/15 all refrigerators
Precision achieved		NP	Precision not reported
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	2	Lompoc completed a refrigerator recycling program evaluation but suffered from a flawed assumption that overstated gross savings and ignored field findings that would have shown considerably less savings. Specifically, the evaluation assumed that all recycled refrigerators would have been used as a second refrigerator. This is not an appropriate assumption, as other scenarios include being kept but not used, discarded to landfill, and discarded to the secondary market where utilization may be less than the entire year.
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	3	The evaluation for Lompoc completed a deemed savings review and adjusted the estimated savings based on the model type and serial number information and obtained energy use from the database of refrigerators maintained by the ENERGY STAR program. The revised savings values were applied to all program participants to derive a total verified savings value for the program.
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	2	Lompoc evaluation provided some discussion of the differences between their verification findings and the ex-ante values, but then suggested that Lompoc use "E3 replacement values" as reasonable although the E3 values generated a program estimate 33% higher than the evaluator's field findings
Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	1	
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NA	
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	NA	
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	NA	
Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	3	
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)	1	No assessment of reliability
<b>Overall score</b>		33	

**Findings and Recommendations-** Since the three programs are relatively similar and utilize similar impact evaluation approaches, this is found to be an appropriate grouping of programs for evaluation.

The key shortcoming of this evaluation is that it did not consider the program participants' intended course of action absent the program or timing of the appliance disposal. The

evaluators implicitly assumed that all recycled refrigerators would be used as second refrigerators (or freezers). This is not usually a good assumption. The evaluation would have benefited from a participant phone interview to better understand participant behavior in the absence of the program (e.g. disposal, sell to secondary market, or use as a secondary appliance).

## City of Palo Alto

**Brief Program Description** – The CPAU residential Smart Energy Program includes a refrigerator/freezer recycling program, and provides incentives for the purchase of the following measures: Refrigerators; Dishwashers; Washing Machines; Gas Furnaces; Gas Boilers; Air conditioners; Water heaters (both standard tank and tankless); Pool pumps, and; Insulation for attic, roof and/or walls.

This is a downstream residential CFL rebate program under the Smart Energy Program. There were 3,908 participants receiving 19,631 CFLs in FY 2007/08. Most of these participants received five CFLs each.

The Commercial Advantage Program provided incentives for the following measures: Commercial Clothes Washer, Centrifugal Water Cooled Chiller, Ceiling Insulation, Unitary System A/C, VFD on HVAC Fan, Window Film, CFLs, LED or Electroluminescent Exit Signs, Occupancy Sensor, T8 Lamps, and Energy Efficient Motors.

The small commercial Right Lights Plus program is a third-party direct install program which offers CFLs, T8 fixtures, LED exit signs, and refrigeration controls and gaskets.

**Brief Evaluation Description** – The impact evaluation confirmed and updated (with DEER 2004) deemed savings values, and reviewed the program database documentation for 224 of the 283 participants (the others were missing records) to confirm that the units were recorded as recycled and that the characteristics of the participating units were comparable with those on which the deemed savings were based.

The impact evaluation verified purchase of CFLs through telephone interviews (conducted during process evaluation) with ~50 participants to achieve 90/10. Interviews did not obtain quantity or wattage of CFLs. Deemed savings were reviewed.

The evaluation of this program was not complete in time for publication with the report. The impact evaluation plans to use multi-variate regression statistically adjusted engineering (SAE) model that uses pre and post program participation billing data along with a number of other potential explanatory variables (ex ante savings and weather data). Will meet or beat the statistical confidence of 90% +/- 10%.

There were 14 unique participating sites in the FY 2007/2008 Commercial Advantage Program. Two were identified as custom measure projects and the remaining a mix of lighting, HVAC, hot water, and Motors/VSDs. Onsites covered six sites accounting for 87% of the claimed electricity savings (~50% from two custom measures, and ~40% from four lighting projects). Exceeded 90% +/- 10% confidence interval statistical guideline.

The impact evaluation stratified savings by measure and magnitude, reviewed engineering assumptions under IPMVP Option A, and verified measures by onsite covering 23 of 66 sites to achieve 90/10. Also, 3 sites with refrigeration measures were certainty-selected, as those measures were seen as likely to grow in importance to the program.



**Table A11 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
				This figure was calculated based on the tables which presented program results. The summary table in the introduction only presented the evaluated programs and not the portfolio.
Gross kWh	NP		4,626,699.00	
Gross kW	NP		NP	
Gross therms	NP		22,131.00	
Net kWh	4,398,899.00		NP	
Net kW	1,950.00		NP	
Net therms	NP		NP	
Lifetime kWh	48,220,815.00		NP	
Lifetime kW	NP		NP	
Lifetime therms	NP		NP	
TRC	2.43		NP	

**Table A12 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments		NP	
Gross kWh	3474663	3,990,672.00	Most likely gross savings
Gross kW	NP	NP	
Gross therms	NP	NP	
Net kWh	NP	NP	
Net kW	NP	NP	
Net therms	NP	NP	
TRC	NP		
Realization rate kWh		115%	
Realization rate kW			
Realization rate therms		NP	

**Table A13 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	3	Comparing SB1037 measure category net savings with Summit Blue's program level gross savings revealed that residential programs accounted for 336,179 kWh of savings more than were reported in SB1037, and commercial programs accounted for 65,200 kWh of savings less than reported, leading to a discrepancy of 270,979 kWh more than SB1037. Summit Blue presented program savings on their own and not in the context of the portfolio, making comparisons to SB1037 on a finer scale impossible.
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	2	Impact evaluation presented results for a subset of programs, not the portfolio.
Percent of portfolio covered (estimated)	%	75%	
Gross kWh of all programs evaluated	kWh	3,474,663	
Net kWh of all programs evaluated	kWh	NP	
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	4	Palo Alto's EM&V report targeted a more complete collection of sites and purposely included the more customized and variable sites within the sample population. Because the report for Palo Alto included a discussion regarding their decisions to study specific programs and sites, appropriate steps were taken to minimize the uncertainty of the findings, although the words "uncertainty" and "risk" were never used.
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	2	
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	3	The Palo Alto report evaluated four programs and utilized the International Performance Measurement and Verification Protocols (IPMVP) Option B for a project involving variable speed drives (VSDs) , which were installed on 12 exhaust and six supply fans for chemical fume hoods. Spot measurements were taken on as many units as possible (some power feeds were too difficult to access), and power meters were installed on five motors of varying sizes for a five week period. This onsite and metering data was used to complete engineering calculations sufficient to meet IPMVP Option B criteria.
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	2	Palo Alto's on-site verification revealed detailed instances of screw-in CFLs being removed prematurely, and stated that this could "severely affect the lifetime of these measures." While the removal of CFLs is arguably more relevant to installation rate (i.e. verification issue), partial credit is given here for mentioning measure life in this context.
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	2	Also 2, NP

Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	3	Also 1, 2, NP
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	The Palo Alto study evaluated a variety of measures using reasonable approaches
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	3	
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	Also NP
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	1	
Precision target			
Precision achieved		NP	
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	3	Also 1, 2, NP
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	3	
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	3	Also NP

Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	2	Evaluation used self-report surveys to explore free-ridership and offered recommendations based on the findings, but failed to develop a net-to-gross ratio (NTGR) or net savings.
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NP	Report used telephone surveys as well, but did not document their sample sizes sufficiently to be evaluated
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	2	The survey posed two questions on free-ridership: (1) whether the participant had considered purchasing the efficient measure earlier, and (2) what the likelihood was (using a five point scale) of purchasing the more efficient equipment absent the program
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	1	The Palo Alto survey asked what the likelihood would have been, absent the program, of purchasing the efficient measures. The question utilized a five point scale to gauge customer responses, which is an appropriate design. However, as the survey did not capture the programs effect on the quantity or timing of the purchase of the efficient measures, it was insufficient to properly account for partial free-ridership

Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	2	Some recommendations provided, but found to be thin (less than 1 page out of a 72 page report). Mostly points out issues, but does provide recommendations for how to address the issues.
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)	1	Report does not mention the reliability of verified savings.
<b>Overall score</b>		48	Individual programs do not rank as high

**Findings and Recommendations-** Summit Blue prioritized limited EM&V funds wisely, alternatively stratifying the evaluation of measures by volume, risk, and future importance to the programs. This allowed the impact evaluations to offer solid recommendations for program improvement on the most critical technologies and practices.

However, several aspects could have been enhanced at little to no additional cost. Most notably, net to gross ratios were not mentioned, and even leveraging typical California IOU evaluation NTGR data could have added significant value. Savings were not specifically referred to as "gross" save for a couple of instances, leading to ambiguity. Demand savings were either omitted, or dispersed and buried in separate tables rather than being compiled in one place like

program energy savings. Similarly, the program savings were presented on their own and not in the context of the portfolio, which would have been easy to do and would have avoided significant confusion. The evaluation of the commercial custom program is written in such an ambiguous fashion that pages 8 and 21 state that there are 14 "unique participating sites" while page 22 states that there are only 6, making it impossible to determine the rigor of the evaluation with any degree of confidence.

## Port of Oakland

**Brief Program Description** – The Non-Residential Custom Program was not described in the evaluation report, but cited as the program that was evaluated. Only one customer participated in the program, and completed a lighting retrofit. Five separate incentive applications were submitted for this lighting retrofit project.

**Brief Evaluation Description** – The evaluation focused on the one site, which included a site visit and review of documentation. A sample was developed for field verification.

**Table A14 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		307,223.00	
Gross kW	NP		66.80	
Gross therms	NP		NA	
Net kWh	279,720.00		NP	
Net kW	61.00		NP	
Net therms	NA		NA	
Lifetime kWh	NP		NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NA	
TRC	1.84		NP	

**Table A 15 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	NP	
Gross kWh	307223	332,837.00	
Gross kW	66.8	77.20	
Gross therms	NA	NA	
Net kWh	NP	NP	
Net kW	NP	NP	
Net therms	NA	NA	
TRC	NP	NP	
Realization rate kWh		1.08	Not reported
Realization rate kW		1.16	Not reported
Realization rate therms		NA	

**Table A 16 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	3	Ex ante Net-to-gross ratios are not round numbers. SB 1037: 279,720 kWh savings. EM&V reported claimed is 307,223 kWh (if this is gross, then 91% NTG applied for SB 1037 net) and EM&V reported verified is 332,837 kWh.
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	3	
Percent of portfolio covered (estimated)	%	100%	One project portfolio
Gross kWh of all programs evaluated	kWh		
Net kWh of all programs evaluated	kWh	NP	Report did not discuss net savings
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	NA	Uncertainty proved to not be applicable because the one site included in the portfolio was verified.
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	3	Small POU that has one project in the portfolio
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	1	On-site count plus wattage adjustments based upon "accepted wattages"
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	EUL not mentioned or addressed in M&E report
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	2	
Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	3	Verification was performed with an on-site count
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	The Port of Oakland study conducted an on-site evaluation, resulting in an adjusted measure count, and adjustments to measure wattage assumptions.
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	2	Early replacement baseline used, no discussion
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	Census of the one project
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	NA	With a census, precision not addressed
Precision target		NA	With a census, target not addressed
Precision achieved		NA	With a census, target not addressed
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	3	
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	3	Port of Oakland, with its one site, was included as having its sampled results applicable to the program as a whole
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	3	Discrepancy was discussed in text to some degree, but not all details were given, a table of calculations would have been helpful, there is no way to check calculations given the limited disclosure
Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	1	No net to gross to identified
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NA	
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	NA	
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	NA	

Section 5 - M&E Summary	Scoring Metric	Score	Notes
Criterion 5.1. Does the evaluation provide clear recommendations?	(1-3)	2	
Criterion 5.2. Does the report assess the reliability of the ex post savings results?	(1-3)		
Overall score		36	

### Findings and Recommendations-

The beginning of the evaluation report cited that five separate incentive applications were submitted, but then it was stated that only one rebate was issued. The report should clarify if the separate incentive applications were combined for one rebate payment.

A sample was developed for field verification, given the number of lighting fixtures retrofitted. The report should describe the sampling plan for the on-site visit, and how many fixtures were counted. Overall, the report would have benefited from a more complete description of the lighting retrofit project, including the baseline fixtures (type and wattage) by space type, along with the retrofit fixture descriptions (type and wattage) by space type. This would have aided with the review of the evaluation report.

The report also would have benefited from tables that compared the application claimed assumptions (e.g. fixture counts, operating hours, wattage) against the verified values.

## City of Redding

**Brief Program Description** – The Redding Electric Utility (REU) Earth Advantage Program is focused on the residential market since this sector comprises 75 percent of REU’s customer base. The REU programs are tracked in multiple databases that are arranged by measure type rather than specific program. Program (portfolio) covers 4 measure categories: HVAC, Energy Star appliances, Lighting and Weatherization

**Brief Evaluation Description** – The impact evaluation focused only on the weather sensitive measures included in the Earth Advantage program. A paper verification was completed for 70 measure installations representing a 90%/+/-10% precision and confidence (no total population count was provided). The verification entailed checking that each installation had a receipt on file, and that the measure listed in the tracking database matched the measure listed on the receipt.

A billing analysis was used estimate the savings associated with four measures: residential HVAC equipment replacement (i.e. air-conditioning), insulation, duct repair and energy efficient windows. A total sample of 1,252 customers was pulled for this analysis, along with the billing data from January 2006 through October 2008. Customers were eliminated if the savings were too small to be seen in the billing analysis, or too large to be realistic.

The analysis resulted in statistically significant realization rates for HVAC and insulation measures, but not for duct sealing and windows. Despite this, the EM&V report combined all four measure realization rates for an overall realization rate of 96% for the group of weather sensitive measures.

Since the evaluation focused on a billing analysis for program participants only, a literature based net-to-gross analysis was completed. The EM&V report included a table of NTG evaluation results for PG&E, CA statewide 2004-05 evaluation, Yolo Energy Efficiency CFL giveaway, Energy Vermont and Energy Trust of Oregon. The results of the literature research were not applied to the verified savings in any way.

**Table A 17 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		3,213,742.00	Unclear if net or gross
Gross kW	NP		1,039.70	Unclear if net or gross
Gross therms	NP		NA	
Net kWh	1,639,577.00		NP	
Net kW	1,233.00		NP	
Net therms	NA		NA	
Lifetime kWh	19,699,232.00		NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NA	
TRC	1.84		NP	



**Table A 18 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	3,315.00	This is for the 4 measure categories
Gross kWh	3213742	NP	No program level ex post provided. Only provided on a per unit of measure basis.
Gross kW	1039.7	NP	
Gross therms	NA	NA	
Net kWh	NP	NP	
Net kW	NP	NP	
Net therms	NA	NA	
TRC	NP	NP	
Realization rate kWh		96%	92% for HVAC, 142% for Insulation
Realization rate kW		NP	
Realization rate therms		NA	

**Table A 19 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	1	EM&V reported portfolio savings seem too high (total of 3,213,742 kWh savings added up across four measure categories mentioned in the EM&V report, compared with 1,639,577 net kWh in SB 1037)
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	2	
Percent of portfolio covered (estimated)	%	75%	Best guess, not clear
Gross kWh of all programs evaluated	kWh	2,420,400	Best guess, not clear
Net kWh of all programs evaluated	kWh	NA	Results were not extrapolated to portfolio
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	2	
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	3	Report looked across portfolio, but subset of measures
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	2	Redding completed a billing analysis that appeared to be in compliance with IPVMP Option C, which is whole facility analysis of utility meter data for the baseline and reporting period.
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	EUL not mentioned or addressed in M&E report
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	2	Redding utilized a billing analysis to evaluate gross measure savings for four of the key program measures. While Redding provided some statistics from their regression models (realization rate coefficient estimates and their associated t-statistics), they did not provide complete model output, which would be helpful in assessing the overall validity of the analysis. Also, while the Redding analysis has the components necessary to provide estimates of total ex post savings, they did not include these estimates in their report, which would have been appropriate.

Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	1	Comparison of hard copy receipts to database
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	The Redding study used a billing analysis to estimate savings impacts of residential HVAC and shell measures.
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	3	
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	2	Redding completed a billing analysis that first segmented the customers that were likely to have a statistically significant result from the billing analysis.
Precision target		90/10	
Precision achieved		NP	Achieved precision not reported
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	3	Redding completed a billing analysis that first segmented the customers that were likely to have a statistically significant result from the billing analysis.
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	1	
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	1	Redding completed a billing analysis, but did not provide any context for how the billing analysis results should be interpreted relative to the original deemed savings estimate. Although the billing analysis results were close to the ex-ante estimate (i.e. realization rate was 96%), a brief discussion of the deemed savings methodology may be helpful to provide insight on why the values are different. Reasons for the difference may also be related to actual cooling degree days relative to the typical meteorological year which is used for building simulations of weather sensitive measure savings

Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	2	The study for Redding offered program design recommendations based on the most recent DEER values for the NTGR of evaluated program measures, but failed to apply the NTGR to gross savings to develop net savings
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NA	The Redding evaluation employed a literature survey to review deemed NTG values, and thus fell into the "not applicable" category with the six remaining studies.
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	NA	Redding employed a literature survey to review deemed NTG values, and thus fell into the "not applicable" category with the six remaining studies.
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	NA	

Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	3	
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)		
<b>Overall score</b>		35	

## Findings and Recommendations-

The report does not provide sufficient contextual information. It provides kWh savings and kW saving across the 4 measure categories shown in the below table. The summation of the kWh and kW savings for these measure categories are far greater than the portfolio savings reported in the SB 1037 annual report.

Within the four measure categories, only HVAC and weatherization specific measures were examined for billing analysis. Although the EM&V report did not specify which weather

sensitive measures from the measure categories were included in the billing analysis, it seems to be the four specific measures with asterisks (\*\*) in the table below. (For example, we presume that “energy efficient windows” matches with “window treatments” as listed in the table below.)

Measure Category	Specific measures installed
HVAC	High efficiency units** Duct repair** Pressure testing HVAC cleaning Swamp coolers Whole house fans Charge test
Weatherization	Insulation** Radiant and thermal barrier Window treatments** Caulking Weatherstripping Water heater wraps
ENERGY STAR Appliances	Windows Refrigerators Clothes washers Dishwashers Water heaters
Lighting	Fixtures, Ballasts Exit Signs Lamps

Realization rates were provided for the four specific measures, but not expanded back to the population. The evaluation could have been improved significantly, if the measure specific verified energy savings were multiplied by the number of units installed to provide a “program” level realization rate and savings value. Furthermore, the savings should be specified as gross or net savings.

Net-to-gross ratio was researched by examining the results of similar programs for five measures: dishwashers, clothes washers, water heaters, screw-in CFLs and windows. These were not the same measures as included in the gross savings estimates based on billing analysis. The literature research did not result in any recommendations for adjusting the NTGR for the impact evaluation. The results were used to inform recommendations on measures to discontinue due to high free-ridership.

Small evaluation budget likely limited the scope and quality of the evaluation. With such a large population of residential participants, the EM&V approach should have considered a telephone free-ridership survey to improve the net-to-gross value.

## City of Roseville (Roseville Electric)

**Brief Program Description** – The Roseville energy efficiency programs described in the evaluation are:

- Residential New Construction Program – This is actually two programs.
  - Preferred Homes, which provides incentives for newly constructed homes to be built beyond Title 24 minimum levels
  - BEST Homes, which provides incentives for newly constructed homes to be built beyond Title 24 minimum levels, but also includes incentives for integrating rooftop solar PV into the new home construction.
- Residential HVAC Program – Incentives are provided for installation of energy efficient air-conditioning and heat pump units in existing homes. Roseville Electric contracts with a third-party to process the rebates for this program.
- Commercial Custom Programs – Incentives are provided for custom projects, which include measures beyond the pre-populated list in the EE Reporting Tool. Roseville Electric pre-inspects the site and hires a third party consultant to review the project for the accuracy of the savings described in the application form.

**Brief Evaluation Description** – Paper and phone verification only. No on-sites, and no re-calculation of any savings (no ex post values provided).

- Residential New Construction Program – The evaluator performed a paper verification of the applications associated with the tracking database. A desk review of the deemed savings methodology was also conducted. No adjustments were made to the deemed savings being claimed, based on improved data or program (site) specific information.
- Residential HVAC Program – The evaluator performed a paper verification of the applications associated with the tracking database. Once again, a desk review of the deemed savings methodology was completed.
- Commercial Custom Programs – The evaluator performed a paper verification of all 21 applications, including the Roseville on-site verification and engineering calculations. One application did not show hours of operation in their analysis, so the evaluator completed a short interview with the customer's engineering staff. Through this interview, the evaluator confirmed that the calculation was correct.

**Table A 20 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh			NP	
Gross kW			NP	
Gross therms			NA	
Net kWh	9,313,572.00		NP	
	2,144 (demand)			
Net kW	2,007 (peak)		NP	
Net therms	NA		NA	
Lifetime kWh			NP	
Lifetime kW			NP	
Lifetime therms			NA	
TRC	3.49		NP	

**Table A 21 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	NP	
Gross kWh	NP	NP	
Gross kW	NP	NP	
Gross therms	NA	NA	
Net kWh	4556296	NP	
Net kW	663.6	NP	
Net therms	NA	NA	
TRC	NP	NP	
Realization rate kWh		NP	
Realization rate kW		NP	
Realization rate therms			

**Table A 22 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	NP	Report does not include any mention of the portfolio level savings.
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	1	Covered three programs. It is not clear what portion of the overall portfolio was included in the three programs. Ex-ante savings were provided for one program – Commercial Custom Impact Evaluation (4,556,296 kWh saved). No ex-ante savings were provided for the other two programs included in the evaluation, so there is no way to estimate what portion of the overall portfolio was covered. Regardless, the evaluation also did not provide an ex-post savings values for any of the programs.
Percent of portfolio covered (estimated)	%	49%	
Gross kWh of all programs evaluated	kWh		
Net kWh of all programs evaluated	kWh	4,556,296	This value represents the Commercial Custom program only (1 of 3 programs evaluated)
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	1	The report looked at where the largest savings were
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	2	
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	1	No savings were re-calculated
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	2	Roseville was found to be a particularly thin evaluation, but the documentation was partially adequate because there did not seem to be much to explain; since no project specific impact analysis was completed, only a simple discussion of the ex-ante savings methodology.

Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	2	Phone verification for custom only, otherwise, paper only
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	1	Report included a review of deemed savings but no site verification. There was some debate as to whether this constituted an impact evaluation at all, or was simply a review of ex ante savings. The report discussed the ex-ante savings methodologies and provided no adjustments based on program reported achievements.
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	3	Title 20 and Title 24 are used as the baseline for Res HVAC and Res New Homes, respectively. Existing baseline was used as the baseline in the commercial custom program (all lighting)
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	Roseville evaluation included a verification of 16 randomly selected projects, which represented the statistical confidence of 90%+/-10%
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	2	
Precision target		90/10	
Precision achieved		NP	
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	1	The Roseville EM&V report reviews the deemed savings approach used by the POU, but provides "findings" such as "we recommend analyzing the updated DEER savings numbers when they are released." In another case, the report reviews the building simulation approach and provides a recommendation that the program account for savings related to surpassing Title 24 code by 20 percent. No revised savings estimates are provided for any specific measures, much less for projects that participated in the program
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	1	
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	NA	Roseville's evaluation report included a short discussion for areas where the deemed savings may be improved, such as updated DEER or PG&E work paper, but no revised values
Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	1	
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NA	
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	NA	
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	NA	
Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	3	For the scope, they did provide recommendations for improving the program
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)		
<b>Overall score</b>		25	

**Findings and Recommendations-** The first and foremost issue associated with this EM&V report is the lack of evaluation of the program savings claims. Evaluator did not place proper emphasis on verifying installation rates or reviewing assumptions associated with the savings calculations, either from secondary literature research or based on actual field observation of equipment performance. Furthermore, an analysis of free-ridership to assess the net-to-gross ratio was lacking.

**Residential New Construction Program** – The evaluator inspected 57 applications out of 315 rebates and stated that the Summit Blue evaluation plan had recommended this number for “90% confidence.” The target level of precision (e.g. 10%, 20%, etc) should also have been reported.

The evaluator mentioned that documentation was inconsistent, but that all documents were located. The evaluator needs to specify what documents were located (e.g. invoices? Application forms?). Furthermore, there was no conclusion whether the program should be awarded full credit or not, for the projects listed in the tracking database. The evaluator should conclude whether an adjustment factor (verification rate) should be applied. Since documentation was inconsistent, but not missing, the evaluator may be able to conclude that 100% of projects in the tracking database should be counted to the program.

A description of the MicroPas software was described as being used for the deemed savings estimate, but there “analysis” did not provide verify any assumptions based on program specific projects. A simple description of the deemed savings calculation algorithm is not sufficient to qualify as an impact evaluation. The evaluator should have at least reviewed the assumptions (e.g. square footage, climate zone, building type) included in the MicroPas building models, and provided improved assumption values.

**Residential HVAC Program** – The evaluator inspected 57 applications out of 350 rebates. One application was found to be missing data associated with the duct test certificate signed by the contractor. The conclusion from the paper verification was that the evaluator had no suggestions for improving the process. This conclusion is found to be extremely lacking. All program processes have room for improvement and the evaluator should also have discussed whether any adjustments should be made to the verified number of completed installations.

A table of the deemed savings associated with different HVAC system sizes and efficiencies was provided. The evaluator states that the DEER savings for these measures are different, with no explanation for why. The evaluator needs to include equations and document the assumptions used by the E3 calculator, compared with the DEER database, and provide an assessment of which savings more accurately represent Roseville’s projects that received incentives. (This assessment may be based on secondary research, or field visits to actual project sites.)

**Commercial Custom Programs** – The evaluator completed a paper verification with one telephone interview. It is recommended that the evaluator at least complete a telephone interview with a sample of customers to confirm assumptions. Even better is for the evaluator to complete on-site inspections to confirm the assumptions.

A more complete description of a site by site evaluation for this program is warranted given the custom nature of the program. A site by site evaluation would include a description of the project, ex ante savings associated with the project, engineering equations and the relevant engineering assumptions (e.g. operating hours, fixture counts, temperature set-point, flow rate, etc).



# Sacramento Municipal Utility District HVAC Evaluation

**Brief Program Description** – SMUD provides rebates and/or SMUD financing for efficiency improvements to homes' building shells and equipment. Improvements include CACs and HPs, duct sealing, windows, attic & wall insulation, insulated siding, solar DHW and cool roofs.

**Brief Evaluation Description** – The evaluation addresses only residential HVAC systems (including ducts). The evaluation involved onsite testing at a sample of participant and nonparticipant homes. Savings were estimated by comparing participant with nonparticipant results. A separate net-to-gross survey estimated a net-to-gross ratio for the program, but these did not inform the net savings, since net impacts were calculated directly. This study did not estimate gross impacts.

**Table A 23 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		NP	The report does not address portfolio savings in any way.
Gross kW	NP		NP	
Gross therms	NA		NP	
Net kWh	84,963,287.00	CY 2006	NP	
Net kW	21,544.00	CY 2006	NP	
Net therms	NA		NP	
Lifetime kWh	1,104,928,081.00	2006 savings for HVAC (cooling+heating+shell)	NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NP	
TRC	NR	Could not find TRCs reported in the FY 05/06 SB1037	NP	

**Table A 24 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	4,424.00	2006 participant population.
Gross kWh	421/427	NP	The study calculates net savings directly by comparing participants and nonparticipants. Because the study also estimates NTGR through a survey, it is possible to back out gross savings, but the estimate would be uncertain.
Gross kW	0.605/0.654	NP	The study calculates net savings directly by comparing participants and nonparticipants. Because the study also estimates NTGR through a survey, it is possible to back out gross savings, but the estimate would be uncertain.
Gross therms	NA	NA	
Net kWh	430/436	502/525	Per unit for Tier 1 and Tier 2, respectively. Report breaks down savings by efficiency/duct leakage/RCA
Net kW	0.633/0.686	0.439/0.459	Per unit for Tier 1 and Tier 2, respectively. Report breaks down savings by efficiency/duct leakage/RCA
Net therms	NA	NA	
TRC	NP	NA	
Realization rate kWh		1.17/1.2	Tier 1 and Tier 2, respectively
Realization rate kW		0.69/0.67	Tier 1 and Tier 2, respectively
Realization rate therms		NA	

**Table A 25 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	1	The evaluation reports ex ante per unit net and gross ex ante savings for the program, but not aggregate program claimed savings. The report included the number of participants for 2006, and we calculated total savings, which were much lower than the 05-06 net savings reported for res HVAC
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	1	The size of the residential HVAC program is not presented in context of the portfolio. No information was provided on why this specific program was selected for evaluation, nor what programs were excluded
Percent of portfolio covered (estimated)	%	6%	Based on data in the 2006 SB1037, residential HVAC is less than 6% of portfolio net savings
Gross kWh of all programs evaluated	kWh	NP	
Net kWh of all programs evaluated	kWh	4,764,344	Res HVAC programs 2006; Estimated from 2006 SB1037
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	1	

Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	1	The evaluation covers only residential HVAC, which is a subset of the Equipment Efficiency Program (as identified in SB1037).
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	3	The SMUD Residential HVAC evaluation was enhanced rigor because data collected included building characteristics, refrigerant charge test, system airflow, power testing, load monitoring, and temperature logging for participating site. Non-participating sites were measured for infiltration, total duct leakage and duct leakage to the outside
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	3	The SMUD HVAC report was a 153 page document that included equations for calculations, telephone survey instruments and on-site data collection instruments.
Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	3	On-site verifications were conducted at a sample of participant homes
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	4	The SMUD HVAC collected detailed on-site measurements at a sample of participant and non-participant homes, and net savings were estimated directly by comparing the results from the two groups.
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	3	The report included analysis and comparison of multiple baselines, with a clear explanation of which baseline was used for the impact estimates and why
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	3	
Precision target		90/12.6	Target defined at 90% confidence
Precision achieved		0.129	
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	3	
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	1	Per unit results are reported but non extrapolated to the program population.
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	3	

Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	5	Net savings were estimated by comparing participant with nonparticipant results. Freeridership of equipment and compliance aspects were quantified by telephone surveys, but not applied to the measurement and testing derived results, for reasons that are explained on page 14 and 43.
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	3	Achieved greater than +/-15% statistical precision at the 90% confidence level. (0.6 cv, sample of 50 for non-participants and 60 for participants.)
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	3	Because net savings were calculated through a direct comparison of participant and non-participant savings, the NTG ratio derived from the free-ridership analysis was NOT used to calculate net savings from gross savings. This is the appropriate methodology to apply, since the measurement and testing portion of the evaluation quantified the savings net of what participants would have achieved without the program.
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	3	Surveys ask questions to reveal partial free-ridership.
Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	3	
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)		
<b>Overall score</b>		51	

## Findings and Recommendations–

This study was the most comprehensive of those examined for this assessment. Nonetheless, there were several ways in which the report could have been improved. The report did not place the evaluated measures in the context of SMUD’s programs and overall portfolio. A cursory comparison of the numbers in the report to the appropriate SB1037 status report raised additional questions of how the evaluated measures fit in, so addressing this would have been extremely useful. All of the results in the study were reported at the per-unit level, and not extrapolated to all participants. It would have been informative to see that extrapolation and a comparison of aggregate ex-ante and ex-post savings. Because the report was so long and complex, it would also have benefited from better organization and additional editing.

# Sacramento Municipal Utility District Refrigerator Recycling Program

**Brief Program Description** – SMUD paid a \$35 incentive to encourage customers to turn in their spare operating refrigerators and freezers. SMUD contracted with a 3rd party to handle pick-ups, transporting and recycling of refrigerators.

**Brief Evaluation Description** – The goal of the study was to determine the number of refrigerators and freezers collected and recycled, estimate average annual kWh and kW reductions per collected appliance and estimate a net-to-gross ratio. The study included in-situ monitoring of refrigerator and freezer energy use, combined with in-situ energy use data collected for another study. In addition to in-situ use, the study looked at at-manufacture energy use and at-death energy use, both based on DOE test procedures. The study included a survey to estimate a net-to-gross ratio.

**Table A 26 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		NP	The report does not address portfolio savings in any way.
Gross kW	NP		NP	
Gross therms	NA		NP	
Net kWh	84,963,287.00	CY 2006	NP	
Net kW	21,544.00	CY 2006	NP	
Net therms	NA		NP	
Lifetime kWh	1,104,928,081.00	2006 savings for HVAC (cooling+heating+shell)	NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NP	
TRC	NR	Could not find TRCs reported in the FY 05/06 SB1037	NP	

**Table A 27 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	8,000	3,538.00	
Gross kWh	1,500	3,874,063.00	
Gross kW	0	501.48	
Gross therms	NA	NA	
Net kWh	NP	2,251,518.00	
Net kW	NP	291.45	
Net therms	NA	NA	
TRC	NP	NP	
Realization rate kWh			
Realization rate kW			
Realization rate therms			

**Table A 28 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	NP	The evaluation reports ex ante per unit net and gross ex ante savings for the program, but not aggregate program claimed savings
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	1	The size of the residential refrigeration program is not presented in context of the portfolio. No information was provided on why this specific program was selected for evaluation, nor what programs were excluded
Percent of portfolio covered (estimated)	%	5%	Based on residential refrigeration net savings from 2006 SB1037
Gross kWh of all programs evaluated	kWh	NP	
Net kWh of all programs evaluated	kWh	4,608,600	Residential Refrigeration Net Savings from 2006 SB1037
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	1	
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	1	The evaluation discusses refrigerator and freezer recycling as a program, while SB1037 characterizes it as a subset of the Residential Appliance Program.
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	2	The SMUD Refrigerator Recycling evaluation report utilized in-situ (i.e. in-house) monitoring of refrigerators and freezers that would be recycled.
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	3	Recycling of refrigerators and freezers in 2006
Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	NP	SMUD refrigerator recycling program evaluation did not report any verification activities. The study appears to have relied on the third-party providers report of refrigerators and freezers picked up.
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	4	The SMUD Refrigerator Recycling program used the results in situ metering done for this evaluation with in-situ metering results from and IOU evaluation and data on at-manufacture energy use to create an accurate estimate of energy savings for the retired appliances.
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	3	The report considered three savings estimates for the recycled equipment: at-manufacture nameplate energy use (DOE test protocol), measured energy use as measured before recycling (DOE test protocol), and measured in-situ energy use (based on monitoring).
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	NP	A small sample of refrigerators and freezers in the program were metered, and that data was combined with in situ metered data from other studies to inform the analysis. Neither the design of the SMUD sample nor the outside sample is discussed in detail.
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	1	
Precision target		NP	
Precision achieved		NP	
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	3	SMUD Refrigerator evaluation included a participant survey to examine what would have happened to the refrigerator or freezer in the absence of the program. This is an important component to examining refrigerator and freezer recycling programs.
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	3	Results are extrapolated to the population of Refrigerator/Freezer Recycling Program participants
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	3	



Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	5	The SMUD Refrigerator Recycling evaluation utilized a self-report survey of program participants to estimate free-ridership. The questions were related to what the participant would have done with the refrigerator or freezer in the absence of the program.
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	3	SMUD's refrigerator recycling study utilized telephone surveys of 203 households and stratified their sample into five categories to determine the NTGR
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	3	The SMUD refrigerator recycling survey had three questions addressing free ridership, including whether the respondent would have kept or gotten rid of the appliance in the absence of the program, and what they would have done with the appliance (stored unplugged, kept it running, gotten rid of by recycling, trading in, selling, etc.)
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	NA	There is no standardized approach for these types of programs.
Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	1	
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)		
<b>Overall score</b>		38	

**Findings and Recommendations-** The study provided a thorough and interesting comparison of refrigerator and freezer energy use at different life stages (at manufacture and at death) and operating conditions (DOE test procedure and in situ). The study provided adequate justification for how it calculated savings. The study did not appear to include verification activities (that is, were the appliances claimed by the program actually picked up and were they recycled, as opposed to being trashed or re-sold). The study did not adequately discuss reliability. A small in-situ metering sample collected for this study was combined with data from another study, and the confidence and precision of neither the individual nor combined samples were discussed. Discussion of reliability for the free-ridership survey was also missing.

## City of Silicon Valley (Silicon Valley Power)

**Brief Program Description** – The Non-Residential Custom Program includes a wide range of measures. The evaluation focused on select projects including s HVAC, lighting, motors/VFS, and “other.”

**Brief Evaluation Description** – A site-by-site evaluation was performed in which claimed measures were verified, possible sources of impact were noted, such as hours of operation, and light, often useless, measurements were taken. No effort was made to extrapolate savings numbers beyond a single site level

**Table A 29 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		24,663,638.00	Only non-res was evaluated. No attempt was made to extrapolate numbers from on-sites to program or portfolio level
Gross kW	NP		NP	Claimed peak kW savings were reported on a site-by-site basis for evaluated sites but nothing was provided regarding whole program or entire portfolio
Gross therms	NA		NA	
Net kWh	24,509,440.00		NP	
Net kW	1,125.00		NP	
Net therms	NA		NA	
Lifetime kWh	376,915,835.00		NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NA	
TRC	4.38		NP	

**Table A 30 – Portfolio Ex Post as Reported by POU EM&V**

Unit accomplishments	NP	6,647,420.00	Non-residential rebated measures only; only total savings for evaluated sites. No aggregated numbers for entire program
Gross kWh	6498237		Provided at site level, but not collected together, nor extrapolated to whole program
Gross kW		NA	
Gross therms	NA	NP	No effort to address net numbers
Net kWh	NP	NP	No effort to address net numbers
Net kW	NP	NA	
Net therms	NA	NP	No effort to address TRC
TRC	NP		Only for non-residential sites that were visited; combines energy realization rate and installation realization rate; includes billing analysis but claims billing proves inconclusive; no effort to address full program population
Realization rate kWh		101%	
Realization rate kW		NA	
Realization rate therms			



**Table A 31 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	3	24,509,440 kWh listed in SB 1037. 24,498,237 kWh listed in EM&V report, but report indicates that site SVP-9 not included in the ex ante because permission could not be obtained to enter the site. The report should state the ex ante savings associated with the missing site SVP-9.
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	3	Report looked at non-residential but does not provide an estimate regarding the percentage of total portfolio kwh savings nor does it provide any program or portfolio-wide estimate of peak kw savings. The report did not provide a percentage of total savings, but was estimated to have captured the majority of savings. 24,663,638 kWh is given as the claimed savings due to the non-residential program. Although the report does not compare this to the overall portfolio, a calculation using the 2006 SB1037 of 24,509,440, would suggest it is over 100%. While this may be comparing gross to net, the report does not approach the topic of net savings, nor does it provide a percentage that is appropriately weighted to net or gross. Despite the lack of an accurate percentage, it appears that a high percentage of SVP's portfolio was evaluated in this report.
Percent of portfolio covered (estimated)	%	101%	
Gross kWh of all programs evaluated	kWh	NP	
Net kWh of all programs evaluated	kWh	NP	
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	2	
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	1	Claims that reported numbers provide analysis for the majority of savings but no actual percentage is defined.
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	2	Silicon Valley Power evaluation took spot measurements of lighting circuits while toggling the time clock to determine which circuit powered a representative number of scheduled lighting fixtures. The evaluation also spot metered site with HVAC measure. For an additional site, the evaluation used HOBO 4-channel loggers to log for one week, and for two more sites, the evaluation monitored AC units with current trend loggers for one to two weeks
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	3	The Silicon Valley Power evaluation included detailed description of the ex-ante savings assumptions and site specific findings, including description of buildings (e.g. vintage), equipment size, operating hours, and parameters metered.

Section 3 - Gross Savings	Scoring Metric	Score	Notes
Criterion 3.1. Does the evaluation include a suitable measure installation verification?	(1-3)	3	
Criterion 3.2. Is the selected savings evaluation method suitable?	(1-3)	3	Billing analysis conducted but was considered inconclusive
Criterion 3.3. Is the choice of baseline suitable?	(1-3)	2	Baseline is explicit for some projects and measures, but not others
Criterion 3.4. Is the sampling approach appropriate?	(1-3)	3	
Criterion 3.5. Were sampling precision targets and achieved precision reported?	(1-3)	2	
Precision target		80/20	80/20 precision discussed but numbers were never extrapolated beyond site-by-site basis
Precision achieved		NP	Achieved precision not reported
Criterion 3.6. Was the selected savings estimation approach executed appropriately?	(1-3)	2	Silicon Valley Power appeared to have the right approach, but rather than developing the more accurate estimates of energy consumption, the report passively identified, but did not resolve the erroneous assumptions held within deemed savings such as incorrect operating hours.
Criterion 3.7. Are the results expanded back to the program population?	(1-3)	1	No extrapolation performed
Criterion 3.8. Does the report explain differences between ex ante and ex post savings results?	(1-3)	2	Silicon Valley Power identified limitations to specific site findings, but claimed incorporating those limitations into an ex-post calculation was beyond the scope of the study. In one site with three 20 hp motors and one 30 hp motor, the assumed annual operating hours of a 20 hp motor were different than those of the 30 hp motor. The reality was that all four pumps were operating on the same schedule. The limited scope of the project was described as the reason that this discrepancy would not be used in the calculation of ex-post savings. The study's scope was acknowledged as the limiting factor that prevented the development of more accurate ex-post savings in four of the ten case studies.
Section 4 - Net Savings	Scoring Metric	Score	Notes
Criterion 4.1. Does the report include a quantitative NTG?	(1-5)	1	No mention of Net
Criterion 4.2. Is the sampling approach appropriate?	(1-3)	NA	
Criterion 4.3. Is the NTG method applied properly?	(1-3)	NA	
Criterion 4.4. Does the approach account for partial free-ridership?	(1-3)	NA	
Section 5 - M&E Summary	Scoring Metric	Score	Notes
Criterion 5.1. Does the evaluation provide clear recommendations?	(1-3)	2	Report provided basic conclusions but left much as inconclusive and what proved conclusive appears questionable in rigor. On the individual site level, the report suggests reasons why their ex-post gross savings differed from claimed savings. No larger effort to provide program-wide recommendations is present
Criterion 5.2. Does the report assess the reliability of the ex post savings results?	(1-3)		
Overall score		36	

## Findings and Recommendations-

Evaluation was more or less a simple verification that identified whether measures were installed. Deemed savings were used along with simplistic survey data, such as occupancy hours, to estimate ex-post gross savings. Basic metering was performed, but often not used

effectively. Ultimately, numbers were not aggregated to the program level, much less the portfolio level. No effort was made to identify net savings, and while a billing analysis was conducted, it was declared inconclusive. With a 101% realization rate for the verified sites, it would be recommended that a more robust metering and modeling effort be conducted to move beyond deemed savings estimates. Furthermore, extrapolating individual sites savings to the program level and providing program-wide conclusions would strengthen the conclusions and broaden scope of the report. Finally, with an analysis of, or at least a justification for the absent analyses of the rest of the SVP portfolio, this report would serve as a more complete evaluation

## City of Truckee (Truckee Donner Public Utility District)

**Brief Program Description** – The evaluation report covered all 17 programs including:

- Residential Lighting Rebate
- Commercial Lighting Rebate
- ENERGY STAR Appliance Rebate
- Electric Water Heater Rebate
- Ground Source Heat Pumps
- Building Envelope & Duct Testing
- Thermally-efficient Windows
- Refrigerator and Freezer Recycling
- Low/Moderate Income Energy Assistance
- Community Outreach & Schools
- Green Partners – Retail
- Green Partners – Restaurant
- Green Partners – Hospitality
- Million CFLs
- LED Holiday Lights
- Low Flow Pre-Rinse Spray Valves
- 2.0 GPM Showerheads

**Brief Evaluation Description** – The evaluation approach included a measure verification based on customer site visits, telephone surveys, billing data, field measurements, lighting logger data and on-site surveys. The report states that IPVMP Options A, B, C and D were used to evaluate energy and peak demand savings for various components of the program offerings. Measurements were short-term and some, but not all parameters were stipulated (as long as the total impact of possible stipulation errors was not significant to the resulting savings). For instance, on-site inspections plus time-of-use adjustments (based on self-report survey) was used to evaluate residential lighting measures. The evaluation included adjustments to EUL and lifecycle savings.

**Table A 32 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		3,910,119.00	
Gross kW	NP		1,059.00	
Gross therms	NP		NA	
Net kWh	4,455,607.00		3,128,095.20	
Net kW	2,705.00		847.20	
Net therms	NA			
Lifetime kWh	41,984,000.00		34,272,223.00	
Lifetime kW	Not Summed		Not Summed	
Lifetime therms	NA		NP	
TRC	7.12		3.44	

**Table A 33 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments		1,282.00	
Gross kWh	1000		
Gross kW	68099.84	52.56	
Gross therms	20.512	NA	
Net kWh	NA	61,023.00	
Net kW	54479.872	42.05	
Net therms	12.8	NA	
TRC	NA	NP	No measure-level TRC provided
Realization rate kWh		1.12	Not reported
Realization rate kW		2.56	Not reported
Realization rate therms		NA	

**Table A 34 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	3	Ex post net savings match claimed. SB 1037 reported 4,455,607 kWh ex ante. EM&V ex ante 3,910,119 kWh. EM&V ex post matched the SB 1037. EM&V reports were completed by February 20, 2009, so in time for the March 15, 2009 report.
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	2	
Percent of portfolio covered (estimated)	%	100%	Extensive EM&V, but evaluation consisted of mostly verification
Gross kWh of all programs evaluated	kWh	3,910,119	
Net kWh of all programs evaluated	kWh	1,059	
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	2	
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	3	All measures/program components evaluated
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	1	On-site inspection plus TOU adjustment based on self report survey
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	3	Truckee Donner, adjusted EUL based on the evaluations findings – the ex-ante EUL was 6.72 and the ex-post EUL was estimated to be 7.27 assuming 8,000 lifecycle operational hours. The EUL was adjusted due to different annual hours of operation. Consequently, the lifecycle savings associated with the installed measure can be estimated for only one of the evaluation reports
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	2	Not sure how delta watts was calculated, assumed deemed
Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	3	On-site count plus adjusted hours of use
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	Adjusted wattage based upon accepted values for replaced fixture - accepted self-reported hour of use
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	2	Early replacement baseline used, no discussion
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	3	223 inspection
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	2	
Precision target		90/10	90/10
Precision achieved		0.0008	Precision of 0.0008 is amazingly low and is suspicious
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	2	The Truckee-Donner evaluation had assessments of 14 portfolio components that were evaluated with varying degrees of rigor. Most of the well-documented elements appear to have been executed properly, but considerable gaps in documentation for other components made it difficult to determine whether the verification and evaluation results was properly calculated. Additionally, the relative precision statistics for the study are suspiciously low
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	2	Truckee-Donner, was given partial credit for insufficient documentation to indicate the expansion methodology. They appeared to have used a simple random sampling where the unweighted realization rate of the sample is applied to the population, but that can't be determined from the report
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	3	Discrepancy was discussed in text to some degree, but not all details were given

Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	5	The study conducted participant surveys to evaluate the net-to-gross ratios (NTGR). The ex ante NTGRs are 0.80. The study conducted participant NTGR surveys and developed specific NTGRs for commercial lighting (0.96), electric water heater rebate (1.0), refrigerator recycling (0.84), Green Partner (0.96), Million CFL (0.90), LED Holiday Lights (0.91), Low-flow Pre-Rinse Spray Valves (1.0), and Low-flow Showerheads (1.0). Otherwise, the study used published values from the EE Reporting Tool and Table 4.2 of the CPUC Energy Efficiency Policy Manual.
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NP	The report used telephone surveys, but did not document their sample sizes sufficiently to be evaluated. There were only 40 surveys, claiming error bound of 0.03, assumed a small CV
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	3	Truckee Donner utilized a net-to-gross participant survey with key questions associated with timing of the program participation relative to the decision-making process. The questions examined instances where participant awareness of the program caused them to purchase and install the efficient measures. Other questions included influence of the program, and self-reported actions that would have been taken in the absence of the program
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	3	Certain aspects of partial free ridership were assessed, and these varied by program.
Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	2	Some Recommendations were provided
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)		
<b>Overall score</b>		49	

## Findings and Recommendations-

Overall, the Truckee-Donner evaluation is found to be relatively complete, with all substantial portions of the portfolio included in the evaluation.

It is unclear what the assumptions were for determining the wattage difference between the retrofit and baseline (pre-retrofit) conditions. Additional data collected associated with room-type adjustment, more surveys, and monitoring of actual consumption could improve the rigor of the evaluation.

Although general equations are provided for the savings calculation, these were mostly associated with the statistical approach to adjusting the program savings. Engineering equations associated with the ex post savings estimates would have helped in the review of the evaluation report. Overall, the achieved precision of 0.0008 is found to be suspiciously low.

## District of Turlock (Turlock Irrigation District)

**Brief Program Description** – Portfolio includes residential rebates, residential audits, refrigerator recycling, shade trees, CFL rebates, new construction rebates, an education program in schools, and an education specialist. On the commercial side are audits, online energy management tools, and custom and prescriptive rebates. The report is unclear about the scope of the evaluation, at various points suggesting that the evaluation covers all non-residential programs and at another referring to just the “Non-residential Custom Program.”

**Brief Evaluation Description** – Impact evaluation utilized site verification and engineering estimates on one large lighting project (warehouses) and one compressed air project (at an industrial food manufacturing facility). The compressed air project included short-term metering. The two projects comprise almost half of portfolio claimed non-residential savings for 2008.

**Table A 35 – Portfolio Ex Ante Savings as Reported by SB1037 and POU EM&V Report**

Portfolio Ex Ante	SB1037 Reported Claimed Savings	Notes	EM&V Reported Claimed Savings	Notes
Gross kWh	NP		NP	
Gross kW	NP		NP	
Gross therms	NA		NA	
Net kWh	10,936,997.00		10,936,997.00	The report identifies these as net savings, but we believe that they may be gross savings
Net kW	1,710.00		1,710.00	The report identifies these as net savings, but we believe that they may be gross savings
Net therms	NA		NA	
Lifetime kWh	125,717,730.00		NP	
Lifetime kW	NP		NP	
Lifetime therms	NA		NA	
TRC	4.53		NP	



**Table A 36 – Portfolio Ex Post as Reported by POU EM&V**

Program	EM&V Reported Ex Ante Savings	EM&V Reported Ex Post Savings	Notes
Unit accomplishments	NP	NP	
Gross kWh	NP	NP	4,853,816 kWh (compared to 4,965,020 kWh ex ante) for the two sites evaluated. The ex post are clearly identified as gross savings; the reported ex ante may be net savings.
Gross kW	NP	NP	440 kW (compared to 447 kW ex ante) for the two sites evaluated. The ex post are clearly identified as gross savings; the reported ex ante may be net savings.
Gross therms	NA	NA	
Net kWh	10936997	NP	
Net kW	1710	NP	
Net therms	NA	NA	
TRC	4.53	NP	
Realization rate kWh		NP	The report claims 0.978 for gross savings for the two sites, assuming that the claimed savings reported in the table are gross savings (ex post savings are clearly specified as gross savings)
Realization rate kW		NP	The report claims 0.984 for gross savings for the two sites, assuming that the claimed savings reported in the table are gross savings (ex post savings are clearly specified as gross savings)
Realization rate therms		NA	

**Table A 37 –Framework Results**

Section 1 - Reporting Adequacy	Scoring Metric	Score	Notes
<b>Criterion 1.1.</b> Does the report include portfolio level ex ante savings that are consistent with SB 1037 reported ex ante savings?	(1-4)	4	
<b>Criterion 1.2.</b> Does the overall evaluation effort adequately cover the POU portfolio of programs?	(1-3)	1	
Percent of portfolio covered (estimated)	%	46%	
Gross kWh of all programs evaluated	kWh	NP	
Net kWh of all programs evaluated	kWh	5,000,000	
<b>Criterion 1.3.</b> Did evaluations consider risk and uncertainty in selecting components to evaluate?	(1-4)	1	
Section 2 - Overview	Scoring Metric	Score	Notes
<b>Criterion 2.1.</b> Does the evaluation report provide an impact evaluation for a single program or the entire portfolio?	(1-3)	NP	
<b>Criterion 2.2.</b> What level of rigor was the evaluation conducted at?	(1-4)	2	
<b>Criterion 2.3.</b> Does the evaluation include EUL and lifetime savings?	(1-3)	1	
<b>Criterion 2.4.</b> Is sufficient documentation provided?	(1-3)	1	

Section 3 - Gross Savings	Scoring Metric	Score	Notes
<b>Criterion 3.1.</b> Does the evaluation include a suitable measure installation verification?	(1-3)	3	On-site verifications were conducted for both sites
<b>Criterion 3.2.</b> Is the selected savings evaluation method suitable?	(1-3)	3	The Turlock Irrigation District evaluated a lighting project using engineering analysis to evaluate a lighting project and short-term metering to evaluate a compressed air project
<b>Criterion 3.3.</b> Is the choice of baseline suitable?	(1-3)	2	The lighting project used the existing equipment as the baseline, which is appropriate for an early replacement situation. However, the remaining useful life of the equipment is not explicitly addressed. Summit Blue references using Title 24 as a baseline
<b>Criterion 3.4.</b> Is the sampling approach appropriate?	(1-3)	1	Turlock was given a "not appropriate," since two sites were evaluated, but no information was provided about the population. The two sites could have represented a census or only a small fraction of the participants. While the sites chosen were large and represent a significant share of both the non-res and overall portfolio, it would be inappropriate to extrapolate from the results to the overall population.
<b>Criterion 3.5.</b> Were sampling precision targets and achieved precision reported?	(1-3)	1	
Precision target		NP	The precision target is not reported
Precision achieved		NP	Achieved precision is not reported
<b>Criterion 3.6.</b> Was the selected savings estimation approach executed appropriately?	(1-3)	2	The Turlock study's evaluation of the compressed air site raised more questions than it resolved. The evaluation identified significant seasonal variation in energy use that the evaluation was not designed to address. Interviews with facility staff were used to seasonally adjust metered energy use from this study and an earlier metering effort. It is unclear how the final verified calculated savings were derived
<b>Criterion 3.7.</b> Are the results expanded back to the program population?	(1-3)	1	
<b>Criterion 3.8.</b> Does the report explain differences between ex ante and ex post savings results?	(1-3)	3	
Section 4 - Net Savings	Scoring Metric	Score	Notes
<b>Criterion 4.1.</b> Does the report include a quantitative NTG?	(1-5)	1	No NTG analysis is presented.
<b>Criterion 4.2.</b> Is the sampling approach appropriate?	(1-3)	NA	
<b>Criterion 4.3.</b> Is the NTG method applied properly?	(1-3)	NA	
<b>Criterion 4.4.</b> Does the approach account for partial free-ridership?	(1-3)	NA	
Section 5 - M&E Summary	Scoring Metric	Score	Notes
<b>Criterion 5.1.</b> Does the evaluation provide clear recommendations?	(1-3)	1	
<b>Criterion 5.2.</b> Does the report assess the reliability of the ex post savings results?	(1-3)		
<b>Overall score</b>		28	

## Findings and Recommendations–

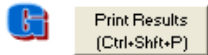
The report is unclear about the scope of the evaluation, at various points suggesting that the evaluation covers all non-residential programs and at another referring to just the “Non-residential Custom Program.” The report does not define or describe the participant population from which the sample of two sites was drawn.

## **APPENDIX B:**

### **Example Guidance for POU's**



Figure 22. Example of the “Results” tab of the 2010 EE Reporting Tool

		Resource Savings Summary									Cost of Efficiency		Cost Test Ratios			
		Units Installed	Gross Annual Energy Savings	Gross Lifecycle Energy Savings	Net Demand Savings (kW)	Net Coincident Peak Savings (kW)	Net Annual Energy Savings (kWh)	Net Lifecycle Energy Savings (kWh)	Net Lifecycle Gas Savings (MMBtu)	Net Lifecycle GHG Reductions	Utility (\$/kWh)	Total Resource (\$/kWh)	PAC	TRC	PCT	RIM
		TOTAL EE PORTFOLIO														
PROGRAM																
CATEGORY	Res Clothes Washers															
	Res Cooling															
	Res Dishwashers															
	Res Electronics															
	Res Heating															
	Res Lighting															
	Res Pool Pump															
	Res Refrigeration															
	Res Shell															
	Res Water Heating															
	Res Comprehensive															
	Non-Res Cooking															
	Non-Res Cooling															
	Non-Res Heating															
	Non-Res Lighting															
	Non-Res Motors															
	Non-Res Pumps															
	Non-Res Refrigeration															
	Non-Res Shell															
	Non-Res Process															